Dear Referee,

thank you very much for your time reading our manuscript and the thorough and helpful corrections and recommendations. Thanks to those, we have improved our manuscripts. Here are some of our additional answers to your requests:

- The reader of the paper easily gets confused with the names of the modelling systems (COSMO, COSMO-CLM, ICON-CLM, ICON-A, ICON-NWP, ICON-LAM, ICON-O, ICON-EUclim, ICON-EU-Nest, ICON-GCM etc). This is in party unavoidable as the names reflect the complex history of the ICON model. Figure 1 is certainly helpful in this respect, but as a reference I would appreciate an additional table that lists the individual configurations of ICON and their basic characteristics. Such a table could also serve as a reference for further publications.

Having a table with individual configurations of ICON is a good idea. However, we would not add an additional table into this paper because it would be a rather complicated table and depart from the main point of introducing and evaluating ICON-CLM here.

- page 1 line 10: "with the setups similar

We changed from setups (plural) to set-up (singular) (page 1, line 11). It fits better in this sentence. In some other parts of the manuscript the plural form setups are still kept after another thought.

- page 2 lines 9-10: It is not really clear to me whether ICON-EU-Nest is a global model with regional refinement over Europe or a higher-resolution version traditionally nested into a global ICON model.

We changed the text. Hopefully it is clearer now (page 2, lines 11-13).

„… ICON-EU-Nest, the regional ICON on European domain interactively nested within the global ICON replaced COSMO-EU (high resolution COSMO model configuration for Europe) for higher-resolution forecasts on the European domain…“

- page 2, line 10: "COSMO-EU" needs an explanation

See previous answer.

- page 2 lines 13-14: As written, the unification was scheduled for end of 2019, which is already in the past. The sentence should hence be modified to "The implies that the last unification of COSMO and COSMO-CLM (COSMO 6), carried out at the end of 2019, was the last one."

The manuscript was written in 2019 but there was a delay in submission till 2020. But the text is also not valid anymore since the plan was changed. The last unification was planned for end of 2020 now. The year was changed accordingly in the text (section 1, page 2, lines 16-17).

- page 2 line 20: Wouldn’t it be better to speak of "climate projection" here instead of"climate prediction" to highlight the longest time horizon for application of the model?

Yes „climate projection“ was added into the text. „Climate prediction“ is still maintained though since they are different (page 2, line 24).

- page 2, line 23: Did ICON-NWP inherit any parts from COSMO (for instance, the physics package)?

The comparison between ICON-CLM and COSMO-CLM is shown in Table 2 and is referred to in Section 3.1. In our opinion, it is not necessary to mention in this part.

- page 2 line 34: Why was CCLM 5.0 clm9 used for the comparison and not the latest(and final) unified version COSMO 6 (see above)?

Because all our work was done in 2018-2019 even before the previous release plan of COSMO 6. The release plan was postponed now till end of 2020. So there is no COSMO 6 yet till this day.

- page 3 lines 13-14: Could you briefly explain why this is the case?

Physically one can feed SST into the regional model also on monthly basis. But we want a flexible option because technically it is easier to update SST at the provided forcing data frequency, so we don't have to prepare specifically the monthly data. The text was re-formulated a bit to make things clearer (page 5, line 1-2).

- page 3 lines 20-21: I suggest to replace this expression by "...time dependent GHGsas provided by corresponding GHG scenarios".

Thanks for the suggestion. We changed in the manuscript (page 5, line 9).

- page 5 line 13: It is not really clear which "necessary changes" are meant here.

The text was changed to: „…After the technical adaptation in the ICON model source code to enable long-term climate simulations…" (page 10, line 16).

- page 7 line 12: How were the data transformed/regridded?

We changed the sentence to „… these data were remapped to the lat-lon grids with the same spatial resolution with the observational data …" (page 9, lines 11-12). Hope it is more understandable now.

- page 7 line 18: Would be better to speak of "ICLM-REF" and "CCLM-REF" here asthe simulations themselves are meant.

Exactly. We changed in the text (page 9, line 20). We also checked again the whole manuscript to make sure that the right names were used.

- page 7 line 28: "a very good performance" -> this statement actually needs somequantification or should, alternatively, be reformulated

We re-structured the manuscript a bit and the sentence was deleted. We also revised all similar statements.

- page 7 lines 28-30: This paragraph actually summarizes the results described later on. It should not precede the presentation of the results I believe, but should rather be moved to the end of the results chapter or even to the conclusions chapter.

The paragraph was moved to the conclusions section. A new small paragraph was written here in this part.

- page 8 line 17: "trends" is misleading here. I'd rather speak of "biases" or "results".

We agreed and changed „trends" to „results" (page 12, line 7).

- page 8 lines 17-18: Any ideas WHY?

It might be due to the positive radiation bias in ICON. We added a discussion into the text (page 12, lines 17-22).

- page 9 line 5: "over the whole evaluation period" -> this expression is misleading as the figure shows no time series of the bias.

Yes, indeed. We changed the text to "averaged over the whole evaluation period" (page 13, line 7).

- page 9 lines 15-16: The numbers obviously refer to events per 20 years. Without providing the length of the period the numbers are, however, not interpretable. I'd suggest to use the unit [days per year] for these numbers and, actually, for the entire Figure 11. This is much easier to understand and to compare to other studies.

Yes that is very true. We changed Figure 11 so that the numbers show days per year instead of per 20 years. The related title and texts were also changed accordingly.

- page 9 line 20: "too low values" -> you're obviously referring to gauge undercatch and evaporative losses here, this should be mentioned (and supported by some reference).

Yes indeed that was what we meant. The text was changed according to your suggestion for better understanding. References for the low precipitation of E-OBS data was also cited (page 13, line 29-32).

- MSLP evaluation in Chapter 4.3: MSLP is evaluated in the same fashion as the other variables, but I see rather little value in this. What is most important here is the spatial MSLP pattern (in addition to the general magnitude), so the evaluation should consider the mean spatial field. The authors might think about replacing their MSLP evaluationby some more informative MSLP analysis.

Thanks for the recommendation. We added an evaluation for MSLP now (Figure 12). The spatial pattern of MSLP from ICLM-REF is compared with the driving data ERA-Interim and additionally with CCLM-REF. A paragraph was also added in the text (page 15 lines 7-17).

- Figure 17: In addition to the naming of the simulations (see above) the variable names in the headers refer to the internal model names of the respective parameters. This is rather cryptic and could be replaced by the full names (2m temperature, cloud coveretc). Alternatively, the abbreviations should be mentioned in the caption.

Perhaps the reviewer referred to Figure 3 (?). We changed the variable names in this figure to full names now. We also revised all naming of the simulations and variables in table, figure, captions, etc. in the manuscript.

- Figure 4: These maps obviously employ some uncommon projection and the Euro-pean continent seems a little "distorted". Why don't you use the same projection as inFigure 2, for instance?

The maps were revised. Hope they look good now.

- Figures 8 and 9: The y-axis of these figures misses a label and the unit of the bias

Sorry it was the result of automatic cropping. All figures should have the full axes now.

- Figure 11: As explained above I'd suggest to use the unit [days per year] or [eventsper year] for these indicators (instead of [days per 20 years]). Furthermore, in topographically structured sub-domains such as AL or SC, the spatial averaging of thenumber of days defined by a temperature threshold makes little sense in my opinion asit completely neglects the large spatial variability. One way to improve on this might beto present the bars as medians with whiskers on top of it reflecting P5 and P95 of thespatial variability within a subdomain. Furthermore, I suggest to place the black EOBSbars to the left of the green simulation bars, not between them (also modify the legendin this case)

Thanks for the comment. We changed the unit from days per 20 years to days per year. However, our main purpose here is to compare the performance of the two models not to look specifically to the spatial variability of the sub-regions. Therefore, we would like to keep the bars as they are. In addition, we also think that adding the whiskers would make the plot very busy and overwhelmed with information. The observation bar we also would like to keep in the middle as it helps with the comparison of both simulations to the reference data.

- Table 1 is hardly readable, the space between the table lines should be enlarged

The space between the table rows were increased, now it is table 2.

SPELLING AND GRAMMAR

Thanks for all of the language corrections. We took over most of them. Only one remark below.

- page 1 line 7: "tests"

- page 2 line 1: "the CLM-Community developed"

- page 3 line 16: "with a user-defined"

- page 3 line 19: "the greenhouse gas"

- page 3 line 23: "retrieve" instead of "get"

- page 4 line 3: "for the European domain"

- page 5 line 14: "was tested"

- page 5 line 33: "Tiedtke/Bechthold" (with a "t")

- page 6 line 34: "British Isles"

- page 7 line 12: "for the purpose"

- Chapter 4 "Results": Past tense is used for describing the simulation results in this chapter. Use of present tense would be more appropriate and clearer in my opinion.

Past tense is actually used in scientific writing in Results section, for example Results indicated that …

Nevertheless, none of the author is native English speaker. We already indicated to the editor that we would agree with a language editing once the manuscript is accepted and in the final form. At this stage there might be more revision, thus it would not bring much to have language editing now.

- page 8 line 12: "...British Isles, Mid-Europe,..."

Dear Referee,

thank you very much for your time reading our manuscript and the thorough and helpful corrections and recommendations. Thanks to those, we have improved our manuscripts. Here are some of our additional answers to your requests:

Main concerns:
1. The manuscript often looks like a technical note of the COSMO community where not everything is understandable for a reader not participating to it. So either this manuscript is a technical note of the COSMO or ICON communities and it does not deserve a GMD publication, or it should be revised to avoid this impression.
Thanks to the comments and recommendations, we improved our manuscript tremendously. Hope it looks better now.
1.1 Not all the terms and projects referred within the manuscript are understandable by a reader not used to the COSMO model. Please consider that most of the readers are not part of the COSMO or ICON community and does not know the related projects.
Ex: COPAT project very often mentioned.
We added in the text the explanation of COPAT project (page 8, line 18), we also tried to explain better other terms. Hope it is clearer now.
Ex: R2B8 configurations and other RxBy.
We tried to explain the RxBy better now in the text. A new section „2. General information on ICON-NWP and ICON-LAM“ was added to the manuscript to describe the ICON icosahedron grid and the meaning of RxBy. A new table (Table 1) was added to give some description of the grids mentioned in the paper.
Ex: Figure 1 is incomplete and not completely self-sufficient. Could you please make the text and the figure 1 consistent: ICON-NWP or ICON-ESM are in the text, not in the figure. Same for the Large-Eddy Simulation version for completeness. Please rework the figure 1.
True that we are still inconsistent in the naming of the different ICON configurations. We did not intend to include the Large-Eddy configuration and the ICON-ESM into the family tree. But after re-consideration, we revised Figure 1, it should be complete now. And the texts were also changed to give a more complete description (page 2, lines 30-32).
1.2 Some figures do not have the quality required for a scientific article. It seems that everything was done a bit too quickly without careful final checking by all co-authors.
Ex: figure 2: missing top line. Please defined the PRUDENCE box abbreviations in the Captions
We changed figure 2 and defined the PRUDENCE box abbreviations in the caption. We also revised all other figures. Please also try to zoom in and out a bit, sometimes the borders of the figures do not show off properly on the pdf viewer.
Ex: figure 3: names of the variables in brackets are not standard names. Please use standard names and consistent naming between figures, tables and text.
We changed the names of the variables in Figure 3 to be consistent with other figures and to be more understandable for the readers.
Ex: many figures without units (fig 4, 5, 8, 9, : : :). Please check
The figures were overcropped. Now they should be find and the units are visible.
Ex: some figures with y-axis labels, other without (cf. Figure 7 vs 8). Please check for consistency.
The figures were overcropped. Now the y axis labels are visible.

2. The manuscript focuses a lot on the comparison with CCLM for the evaluation run. I’m not sure that this should be the major point of such an article.
2.1 I understand the will of the authors to focus on this comparison but I would expect that the reader is likely more interested by a fine description of the behaviour of ICLM itself. So at least in the section 4, please spend more time in describing the ICLM behaviour and less time to the description of the CCLM behaviour. You may even want to cut section 4 in two parts, one dedicated to the new model evaluation and the other to the quick comparison with CCLM.
Thanks a lot for this suggestion, we splitted Section 5 into different sub-sections, one is Technical tests and others for Evaluation and comparison with COSMO-CLM.

2.2 In addition, personally I would find more relevant to compare ICLM with all the EURO-CORDEX evaluation simulations performed at 12 km. You can either use the results obtained in Kotlarski et al. (2014) if values are numerically available or recompute some of the key scores using data downloaded from the ESGF. Doing so, you will place the newly-developed RCM within the state-of-the-art of the RCMs in Europe. I know that this request requires massive additional work but I hope that the authors will consider it.
Kotlarski, S., Keuler, K., Christensen, O. B., Colette, A., Déqué, M., Gobiet, A., ... & Nikulin, G. (2014). Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble. Geoscientific Model Development, 7, 1297-1333.
We agreed with the reviewer that comparing ICLM-REF to the EURO-CORDEX simulations is very tempting. However, we do not consider it a must. Taking this publication below for instance:
GIORGETTA, Marco A., et al. ICON-A, the atmosphere component of the ICON Earth System Model: I. Model description. Journal of Advances in Modeling Earth Systems, 2018, 10. Jg., Nr. 7, S. 1613-1637.
They compared their newly introduced model to a precedent model, in their case ECHAM, and not to many others.
As ICON-CLM is not yet well tuned, it would not be really fair to compare ICLM-REF to the CORDEX ensemble. Nevertheless, we agreed to add a comparison using the figures in Kotlarski et al. (2014). A description was added in the Evaluation methods section (page 10, lines 1-13) and a discussion of the results on page 11, line 28 to page 12 line 6.
We plan a more detailed comparision with other models once the model is fully tuned.

2.3 Reading the text, I often feel that the authors are “too proud” of ICLM being so close or even better than CCLM. Again I understand the author point of view after so many years of work and the fear of not being as good as the old model. However the way it is phrased is not scientific (objective) enough and show too much satisfaction with themselves. See for example, conclusion, abstract, page 7 line 28-30. Please re-read the whole manuscript and rephrase keeping in mind that the goal of the paper is to present a first version of the model and not to “kill” the old one. Model developers are often not well placed to judge themselves their new model. At the end, it will be up to the readers and then to the ICLM users to decide if the new model better fit their applications. The future will tell us.
It has never been the authors‘ intention to be too proud or to kill the old model. We re-read the manuscript again and still felt that the wording is quite appropriate, perhaps it is just a habit of language. However, we still did some tone down at some places. Hope that it sounds better now.

3. In a first paper describing a new RCM, I m expecting much more illustrations concerning the technical tests performed with the model before the evaluation run. Many tests are mentioned (time steps, domain decomposition, different computing system) but not really exploited and illustrated. Even if those tests are very appealing to present, for me, they should be at the heart of such paper and each test should be documented by a table or a figure. Currently we need to trust the authors blindly concerning the test results without any proof or trace.
The different time step tests were illustrated in Figure 3 and discussed in the text.
We added a description of the domain decompostion test now to the text (page 10 lines 18 to 23). But showing figure or table for these tests is difficult since the results are binary identical. In our opinion, the current text clarifies what we tested and how the results are.
An explanation of the different computing system test was also added to the text (page 11, lines 1-5).

3.1 I’m advising to create a section dedicated to the model tests. That is to say to split section 3.1 in two sub-sections, one describing the tests and one for the evaluation run.
3.2 I’m also advising to add at least the “1+1=2” test. That is to say, checking if running 2 months in one job or in two jobs with a restart between the months give the same results or not. This allows to verify the restarting procedure.
Thanks for this interesting advice. We performed two additional tests according to the referee’s recommendation: (1) 2 months in one job without a restart; (2) 2 months in 2 jobs with a restart.

We compared the result of this new tests. The results were identical. A description was added in the text (page 10, line 29-32).

3.3 Later (not for this specific article), I'm also advising to test the model in the Big-Brother / Little-Brother framework what is for me a mandatory step for any new RCM (see for example Denis et al. 2002)

Denis, B., Laprise, R., Caya, D., & Côté, J. (2002). Downscaling ability of one-way nested regional climate models: the Big-Brother Experiment. Climate Dynamics, 18(8), 627-646.
Thanks again for the interesting idea. We will consider that for our next steps with ICON-CLM.

4. Not enough information on the model configuration and simulation setup. In such article, I'm expecting more information about the model itself and its configuration for the evaluation run. The information given in section 2.1 and in section 3.1 are not complete for me.
4.1 First, clarify what should fit in section 2.1 and what should fit in section 3.1. For me everything general concerning the model itself should go in 2.1 whereas the specific model setup for the simulation (domain, resolution, time step, physical choice, tuning, forcing choice) should go in 3.1. The separation is not always easy but deserve some attention to ease the reading.
We looked at section 2.1 and 3.1 again, in our opinion these two sections are already separated, there is no information about the model setup in section 2.1 and vice versa there is no information about the model itself in section 3.1.

4.2 For the model description (section 2.1), I'm expecting more information and related tables and figures on the horizontal grid (how does the icosahedric grid look like ?), the distribution of the vertical levels, the output procedure (do you output on the icosahedric grid or on a more classical grid ? See the text page 7, line 11). Do you have the option of spectral nudging in addition to the upper boundary nudging? Also add more information about the relaxation zone and lateral nudging procedure (width, variable nudged, strength of the nudging, filtering tricks if any: : :) for example in the paragraph page 4 line 12-16.
We added Table 1 to give some charateristics of the grids that were used and referred to in the paper. A description of the icosahadric grid was added into section 2 (page 3, lines 17-27). We also added Figure 2b in addition to Figure 2a to show the triangular grid R2B8. With regards to the vertical level distribution a paragraph was added in section 2 (page 3 line 28 to page 4 line 5). We output on the rotated lat-lon grid not the icosahedric grid, but it can be an option in ICON. This information is added in section 2 (page 4, lines 6-10) and section 4.2 (page 9, line 10).
In ICON limited mode, there are only options for global data nudging or the vertical velocity with the damp layer is damped towards zero. This was already explained in section 2 (page 4, lines 18-22).
The information about lateral nudging was added in section 2 (page 4, lines 11-17).

4.3 For the simulation setup, I'm expecting there the number of grid meshes for the EURO-CORDEX configuration, the way to define the grid, the numerical cost (compared to CCLM at least), the resolution (explain what R2B8 is) but also the description of the forcings of the run. In particular, in addition to the GHG, SST and sea-ice cover (described in section 2.1), I'm expecting some information concerning the aerosol representation (3D+time variation) that can be very variable from one RCM to another (Gutiérrez et al. 2020), the tropospheric ozone and the evolution of the land-use-land-cover if any (Davin et al. 2020).
Gutiérrez C., Somot S., Nabat P., Mallet M., Corre L., van Meijgaard E., Perpiñán O., Gaertner M.A. (2020) Future evolution of surface solar radiation and photovoltaic potential in Europe: investigating the role of aerosols. Environ. Res. Lett.,15 (3), 034035, https://doi.org/10.1088/1748-9326/ab6666
Davin, E. L., Rechid, D., Breil, M., Cardoso, R. M., Coppola, E., Hoffmann, P., ... & Raffa, M. (2020). Biogeophysical impacts of forestation in Europe: first results from the LUCAS (Land Use and Climate Across Scales) regional climate model intercomparison.

Earth System Dynamics, 11(1), 183-200.
The number of grid meshes for the EURO-CORDEX configuration was added in Table 1. The meaning of R2B8 and general information about the ICON grid was added in section 2, page 3, lines 17-27. The performance in terms of speed with comparison to COSMO-CLM was added in section 5.1, page 11, line 4-5. We have not optimized the run configuration for ICON-CLM, therefore, we would not further comment on the computing cost at this stage. We expect that after an optimization ICON-CLM computational cost would be much less than at the moment.
Information about the aerosol and ozone climatology we used for our simulations was added in section 4.1, page 7, line 14-15.

5. A tricky point in RCMs is the capacity to keep or to modify the large-scale information provided by the driving model. Many methods can be applied to check this (Big-Brother/Little-Brother experiment, see above or GCM-RCM temporal or spatiotemporal correlations for large-scale fields often in altitude or cyclone tracking or weather regimes identification). You may want to keep it simple for this study but could you please show at least one illustration allowing to check the lateral forcing procedure? For example, you may want to correlate the Z500 anomaly or the temperature in altitude between the model run and the driver (ERA-Int) at various temporal scales (e.g. yearly, seasonal, monthly, daily, 6-hourly) or anything showing to the reader that ICLM is able to reproduce the large-scale of the driving model at least for some temporal scale (see for example Sanchez-Gomez et al. 2009).
Sanchez-Gomez, E., Somot, S., & Déqué, M. (2009). Ability of an ensemble of regional climate models to reproduce weather regimes over Europe-Atlantic during the period 1961–2000. Climate Dynamics, 33(5), 723-736.
We calculated the correlation of the geopotential at 500 hPa between ICLM-REF and ERA-Interim data for different time scales (6 hourly, daily, monthly, seasonal, yearly) as suggested. The results are shown in figure 13, a discussion was added in the text as well (section 5.3). In a word, the correlation is pretty high, averaged values higher than 0.925 for all time scales. Correlation is better with longer time scale. ICON-CLM seems to not distort the large scale information of the driving data.

6. Minor comments:
6.1 page 2, line 24: could you explain the difference between “one-way nested subdomain” and “limited-area mode”? For me, it is the same thing. Is it a question of on-line versus off-line?
Yes correct, that is the difference between “one-way nested subdomain” and “limited-area mode”. With “one-way nested subdomain” the nested subdomain and the global domain are being simulated at the same time. Global domain gives forcing to the nested domain, but there is no feedback from nested domain back to global domain.
With “limited-area mode”, there is no global model. The boundaries are simply prescribed from external data. We find that the texts are clear enough, and that the referee could already understand the difference, and plus the “one-way nested subdomain” is not the focal point here, we did not change the text.

6.2 page 3, line 16: for the update of the SST, could we also use lower frequency such as daily or monthly?
Yes physically one can feed SST into the regional model also on monthly basis. But we want a flexible option because technically it is easier to update SST at the given forcing data frequency, so we don‘t have to prepare specifically the monthly data. The text was re-formulated a bit to make things clearer (page 5, line 1-2).

6.2 page 3, line 19: green house ! greenhouse
We changed from green house to greenhouse in the text.

6.3 page 4, line 6-11: this paragraph could perhaps include more information about the input/output procedure, the file format, the flexibility of the outputs, : : : For example, is it possible to output hourly precipitation and monthly-mean MSLP from the same run? or do you need to output all variables at the same frequency before a post-processing step?
Yes, it is possible to write out different variables with different temporal resolution we added this information in Section 2, page 4, lines 7-10.

6.4 page 4, line 33: grammatical issue
We re-read the line but did not find any grammatical issue. Please re-consider this comment or make it clearer.

6.5 page 4, line 24-30: could you explain more the restart procedure and the job management and its flexibilit? Could you perform daily run, monthly run, yearly runs ? Or do you have a mandatory time slice such as one month?
Yes one can run the model for a wished time period (that is why we could be able to do the 1+1=2 test with two months in one job without a restart). But we normally choose calendar month. Thanks. We added this information to Section 3.2, page 6, lines 17-19.

6.6 page 5, line 2: could you tell more about the tuning strategy for ICLM. What do you try to optimize?
What we meant with these text is to introduce the Starter Package of ICON-CLM as an useful tool for different purposes. One is using the Starter Package in tuning ICON-CLM. It is not our intention to tune or to set the strategy of tuning ICON-CLM.
This work is planned in the next phase of COPAT project and will be introduced later.

6.7 page 5 and in many places: EU-CORDEX ! EURO-CORDEX
We changed in all places to EURO-CORDEX.

6.8 page 5, line 20: 30 km. Give also the value in hPa.
Yes we gave now the value in hPa (page 7, line 11)

6.9 page 5, line 29: give the list of the variables nudged and the nudging coefficient
We added this information on Section 4.1, page 7, lines 22-23.

6.10 page 5, line 31-33: The use of many unexplained grid names (R2B8, R3B8, R3B7) is confusing. Simplified or explain. Also in the paragraph, you mention tuning parameters from global settings but setup from LAM: : : clarify
We added the description of the ICON grid in which the names and denotes of the grids are explained in Section 2. Also Table 1 gives information on the mentioned grids.

6.11 page 6, line 8: could you compare the 120 s time step with state-of-the-art RCM time steps at the same resolution?
We do not have information about the time steps from other RCMs, they are also not stated in Kotlarski et al. (2014). Time step of COSMO-CLM at 12 km is 100 s, this information is added on page 8, line 12.

6.12 page 20-23: is the reference CCLM simulation published? Any reference to refer to? If yes, cite it. If not, you need to describe it in the method section or to use a published run such as one of the EURO-CORDEX evaluation simulations performed with CCLM and available on the ESGF.
The run CCLM-REF is unfortunately not published yet. But it was done with the most recently recommended version and configuration of COSMO-CLM. That's why we chosed this run and not one in EURO-CORDEX evaluation simulations.
The description of the CCLM-REF is in section 4.1, page 7, line 15-23.

6.13 page 6, line 34: clarify that you are considering only land points.
Yes we clarified this in Section 4.2, page 9, line 14.

6.14 page 7, line 12: typing issue?
Yes indeed. The letter „f" is missing in front of the „or". We corrected now.

6.15 page 7, line 21: In your case, if I understand well, the RMSE measures a skill related to temporal variations of the variables over the PRUDENCE boxes. So I would have dedicated STDEV to a spatial skill score by averaging in time before computing the standard deviation. Currently STDEV is spatio-temporal score if I understand well, what is therefore quite difficult to interpret. Please, consider to change this. Also table caption mentions "spatial standard deviation" whereas the text mention "spatiotemporal

standard deviation”. Please clarify.
Yes, the sdtdev was spatio-temporal score. Thanks for the advice, we changed it to spatial by calculating the time average and then the deviation. The text was also changed to „spatial“. The description of the STDEV was adapted (Section 4.2, page 9, line 29-33). The discussion of the results was also changed accordingly.

6.16 page 7, line 21: For the quantitative score, I’m not forcing you to do so but it could have been a better option to plot Taylor diagrams (incl. RMSE, correlation, standard deviation) in order to be more exhaustive in the evaluation of the runs: for example a spatial Taylor diagram per season for all European land points and a temporal Taylor diagram for each PRUDENCE box. This is just an advise. In particular, it allows to put all boxes or all seasons or all variables on the same figure.
We did actually make the Taylor plots at the beginning, simply because they are part of the Evaluation tool of ICON-CLM and are made automatically when the tool is run. But then we decided to present our results in the form of the tables. Perhaps it is a matter of choice.

6.17 page 7, line 21: If you decide to keep the score STDEV, I propose to put in the tables the ratio of the standard deviations (Model/Obs) in order to have only 2 columns as for the RMSE allowing to easily see the best model for every line.
We changed the stdev according to the referee’s comment.

6.18 page 7, line 28-30: this small paragraph illustrates well my major comment 2.3 with terms such as “very good performance”, “consistent for all six evaluated variables”, “already of similar”. Please rephrase in a more objective and scientific way without overstating the results obtained. Also remember that the ICON project started 20 years ago. So the model is not so new and has been already tuned and adapted at that resolution over the European domain. I’m aware that a model used in climate mode can show biases not seen in weather forecasting mode but still, you are building on the weather forecast experience. Also note that the model performance is not “consistent” for all variables. From my point of view, it seems better for temperature-related variables than for precipitation or MSLP. Here again, a section comparing the ICON run with all the Euro-CORDEX RCM runs in evaluation mode would be more conclusive (see previous major comment).
The mentioned small paragraph was removed and replaced.

6.19 page 8-9-10: Please reorganise the text of those sections to put first the description and discussion of the ICON biases before comparing more quickly with the CCLM reference as the reader want more information about the strengths and weaknesses of ICLM and less about CCLM. Currently I find that the ICLM description is too light and the CCLM description too fat.
The manuscript was changed heavily. We hope the ICLM part is better now.

6.20 page 8, line 10: “no bias” ! When the median bias is near zero, it does not necessarily mean “no bias”, it can mean “bias compensation in space”. Rephrase.
In the text, we wrote „nearly no bias“ as the medians were about 0.01 K and the percentile boxes were pretty short.
We re-formulated to „relatively small bias“ now (page 11, line 21).

6.21 page 8, line 17: “extreme daily temperature” ! avoid to use the world extreme for min and max daily temperature. It is misleading for the reader as “extreme” is often kept for specific statistics or indices. Check everywhere. Also page 9, line 2, line 3.
Thanks for the comment. We checked everywhere in the manuscript and replaced the word „extreme“ with other expression.

6.22 page 8, line 21: “the bias was larger”. All the text of the results is written at the past form. I’m not an English specialist but it would be easier to read at the present form ! “the bias is larger ...”. Please consider to change this everywhere in the results section.

The authors are also not English specialists. But from the teaching of tense in academic writting, for results section, past tense is used to describe results obtained. Simple present is used to describe figures, tables. So we would like to keep the tense as it is for the moment.
The manuscript will go through English editting once it is accepted.

6.23 page 9: please state that ICON is not so good for Summer day statistics. I don't understand why the representation of figure 11 is not similar to the representation of figures 5 to 10 with a box plot representation. A black box can be used for the observation in addition to the green and blue boxes in that case.
Yes we added in the text that ICLM-REF is not as good for summer day as for other temperature-related indices. But still it is beter than CCLM-REF in 3/8 sub-regions (ME, SC, AL) and on average over the whole Europe (EU). In three sub-regions, both experiments were equal (BI, IP, FR).
In all figures from 5 to 11, blue is ICLM-REF, green is CCLM-REF, and in figure 11 black is observation. In figure 5 to 10, there is no black because they show already the biases. We could add black boxes representing observation in figures 5 to 11, but since each sub-figure has already 16 boxes, we do not want to make the figures too crowded. We would like to keep figure 5 to 10 as they are, with the biases.

6.24 page 9, line 19: not sure I agree that CCLM overestimates the precipitation. It is relatively well balanced over Europe contrary to ICON.
As stated throughout the manuscript, CCLM-REF was better for precipitation. But still it simulated more precipitation in comparison with E-OBS (of course E-OBS properly has measuring error, see the next comment). This overestimation can be seen in figure 4, and even more obvious in figure 11.

6.25 page 9, line 20. Please cite a reference for the "too low values". For precipitation, please also mention and discuss the strong model biases over the topography.
We added a citation for the low values due to gauge undercatch and evaporation. The text was also revised a bit too make it clearer what we meant (Section 5.2.2, page 13, lines 30-32).

6.26 page 9, line 29: "summer had the smallest variations". Not so true if you think that precipitation is very low in summer for some regions. Computing the error in % (even without showing them) may help for discussing the results
That is true that due to the low precipitation values in summer, the variation of the values are lower than in other seasons. But the statement about the bias was correct.

6.27 page 9, line 34: "five out of height" ! for me it is 7 out of 8. Please check in table 6.
The stdev was replaced by the stdev ratio (model/obs). The text was changed accordingly, it is six out of eight.

6.28 page 10, line 3-6: for me by eye, CCLM-REF seems better than ICLM-REF for those indices. Please re-assess.
For Wet days index ICLM-REF was better in 6/9. Heavy precipitation 4/9, very heavy precipitation 3/9. That makes 13/27, and few times the 2 models are give almost the same results. We would keep the same statement that none of the models is better than the other for these indices.

6.29 page 10, line 7: please cut the MSLP and cloud section in two sections, one for each variable for consistency and add the Table 8 for the cloud cover again for consistency.
We cut the MSLP and cloud into two sections (Section 5.2.3, 5.2.4) and table 9 was added for cloud cover.

6.30 page 10, line 8-11: any explanation for the MSLP biases in both models?
We do not know. ICLM-REF has a bit higher MSLP than CCLM-REF which shows in its positive MSLP biases, compared to the negative biases from CCLM-REF. The pressure pattern from the driving model is quite well kept in both simulations though.
An additional evaluation for MSLP was added (Figure 12). In this figure, we compared MSLP from ICLM-REF and CCLM-REF to that of the driving model ERA-Interim.

6.31 page 10, line 9: for MSLP, it seems that the biases can reach values higher than 2.5 hPa (cf. Figure 4 over Spain for ICLM.

It is true. The text was changed in page 14, line 28.

6.32 page 10: same question for the cloud biases. Any explanation or hypothesis?
We just know from ICON that it produces a bit too little cloud and therefore has a positive bias in radiation. Perhaps this underestimation of cloud is stronger in COSMO-CLM. But we do not know for sure about COSMO-CLM, therefore would not add further comment on this.

6.33 page 10, line 20: not clear where you find the +/- 5% values
This line discusses bias of cloud cover in Figure 4. On ICLM-REF side the colors are mostly light pink/blue which correspond to -0.05 to 0.05 and translate to +/- 5%. We added the explanation in the text to make it clearer (Section 5.2.4, page 15, line 21).

6.34 page 10, line 20-21: "overestimation of the cloud cover : : : cold bias". ok for the causality for tasmax but this is often the opposite for tasmin. Rephrase.
What is refered to here is tas. Looking at Figure 4, the overestimation of cloud cover aligns with where CCLM-REF has cold bias for 2 m temperature. We added the reference to Figure 4 to the text to avoid confusion (Section 5.2.4, page 15, line 23).

6.35 page 11, line 11: Personally my assessment is that CCLM is better than ICLM for precipitation. Please re-assess. I agree that models are equivalent for MSLP.
It is hard to say from the areal average since some part ICLM-REF is better and some part CCLM-REF is better. But we agree that overall CCLM-REF is better with precipitation and editted the text (Section 6, page 17, line 4)

6.36 figure 3: please make this figure easier to read. For example by increasing the thickness of the curves? Possibly showing only seasons or showing maps? Try to make it simpler and more informative with the key message easier to catch for the reader.
Our idea with this Figure 3 is to test whether ICON-CLM gives remarkably different results due to the choice of time step, which happened with COSMO-CLM. The figure shows that the lines are quite close to the others and no line stands out of the bunch. We would like to keep this figure like that. Of course as written above, we changed the name of the variables and made the sub-figures a bit nicer now.

6.37 figure 4: showing the areas where the differences are statistically significant or not may lead to a more informative figure and make it easier to describe in the text in order to focus only on signaificant biases. Please revise the map projection for figure 4 (it is ugly currently) in order to limit the zone without information in each panel. Also "shave your model" that is to say remove the relaxation zone or comment the model behaviour there in the text.
We revised the map projection and removed the relaxation zone.

6.38 figure 5-10: I like such figures. Check the y-axis labels and the units everywhere.
Yes the figures were overcropped. Now y axis labels and units are visible.

6.39 Table 1: Please add more information about the physics by splitting the deep convection and shallow convection lines and by splitting the radiation in short-wave and long-wave radiation. Add in this table all useful information and references for the physics as it will likely serve as reference for many articles afterwards.
We added more information regarding shallow/deep convection and short/long wave radiation as requested.

6.40 Table 8: please add a table for the cloud cover

We did. It is table 9.

ICON in Climate Limited-area Mode (ICON Release Version 2.6.1): a new regional climate model

Trang Van Pham[1], Christian Steger[1], Burkhardt Rockel[2], Klaus Keuler[3], Ingo Kirchner[4], Mariano Mertens[5], Daniel Rieger[1], Günther Zängl[1], and Barbara Früh[1]

[1]Deutscher Wetterdienst, Frankfurterstr. 135, 63067 Offenbach am Main, Germany
[2]Helmholtz-Zentrum Geesthacht, Max-Planck-Str. 1, 21502 Geesthacht, Germany
[3]Brandenburg University of Technology, P.O. 10 13 44, 03013 Cottbus, Germany
[4]Freie Universität Berlin, Carl-Heinrich-Becker-Weg 6-10, 12165 Berlin, Germany
[5]Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

Correspondence: Trang Van Pham (trang.pham-van@dwd.de)

Abstract. For the first time the limited-area mode of the new weather and climate model ICON has been used for a continuous long-term regional climate simulation over Europe. ~~Building upon~~ Built upon the limited-area mode of ICON (ICON-LAM), ICON-CLM (ICON in Climate Limited-area Mode, hereafter ICON-CLM, available in ICON Release Version 2.6.1) is an adaptation for climate applications. A first version of ICON-CLM is now available and has already been integrated into a starter package (ICON-CLM_SP Version Beta1). The starter package provides users with a technical infrastructure that facilitates long-term simulations as well as model evaluation and test routines. ICON-CLM and ICON-CLM_SP were successfully installed and tested on two different computing systems. ~~Test~~ Tests with different domain decompositions showed bit-identical results, and no systematic outstanding differences were found in the results with different model time steps. ICON-CLM was also able to reproduce the large-scale atmospheric information from the global driving model. Comparison was done between ICON-CLM and COSMO-CLM (the recommended model configuration by the CLM-Community) performance. For that, an evaluation run of ICON-CLM with ERA-Interim boundary conditions was carried out with the ~~setups~~ set-up similar to the COSMO-CLM recommended optimal ~~setups~~set-up. ICON-CLM results showed biases in the same range as those of COSMO-CLM for all evaluated surface variables. ~~This is remarkable because the~~ While this COSMO-CLM simulation was carried out with the latest model version which has been developed ~~for two decades~~ and was carefully tuned for climate simulations on the European domain~~. Furthermore~~, ICON-CLM ~~already~~ was not tuned yet. Nevertheless, ICON-CLM showed a better performance for air temperature, its daily extremes, and slightly better for total cloud cover. ~~Results for~~ For precipitation and mean sea level pressure~~did not show clear advantage from any model~~, COSMO-CLM were closer to observations than ICON-CLM. However, as ICON-CLM is still in the early stage of development, there is still much room for improvement.

1 Background information

In 1999, the limited-area weather forecast model LM (Lokalmodell, Doms and Schättler (1999), later COSMO, Baldauf et al. (2011)), which was developed by the Deutscher Wetterdienst (DWD, the German Meteorological Service), went operational together with the global model GME (Majewski and Ritter, 2002). A few years later, it was renamed into "COSMO model" in order to reflect that further development has become a joint task of the COnsortium for Small scale MOdelling (COSMO). In 2002, the Climate Limited-area Modeling Community (CLM-Community) developed the first version of the regional climate model named CLM. In 2007, the developments in COSMO and CLM were recombined and a first unified version of the weather forecast and climate modes, named COSMO-CLM (Rockel et al., 2008), was released.

In 2001, a cooperation between DWD and Max-Planck Institute for Meteorology (MPI-M) was initiated, with the aim to develop a new modelling system for weather forecast and climate prediction. The new system was intended to replace the existing system COSMO/GME for operational weather forecast on one side and, on the other side, the global climate and earth system model ECHAM6/MPI-ESM (Stevens et al., 2013; Giorgetta et al., 2013). As a result of this initiative, the global numerical weather forecast model ICON (Icosahedral Nonhydrostatic) (Zängl et al., 2015) was developed and replaced GME as the operational model at DWD on the 20th of January 2015. As a next step, in December 2016, ~~a domain with grid refinement over Europe (~~ICON-EU-Nest~~)~~, the regional ICON on the European domain interactively nested within the global ICON~~replaced the~~, replaced COSMO-EU ~~for~~ (high resolution COSMO model configuration for Europe) for higher-resolution forecasts on the European domain. In the second half of 2020, the convection permitting configuration of ICON-LAM (ICON-D2) became pre-operational. According to the plans, ICON-D2 will replace the high resolution COSMO-D2 for the German domain early 2021 and DWD will stop the operational use of the COSMO model after more than 20 years. This implies that the next unification of COSMO and COSMO-CLM (COSMO 6), scheduled for the end of ~~2019~~2020, will be the last one. Afterwards, the support for COSMO and COSMO-CLM will be gradually reduced.

In this work, we prepared state-of-the-art tools for climate applications for the upcoming years. Starting in 2017, DWD and the CLM-Community decided to develop a new regional climate model (ICON-CLM) based on the Limited-Area Mode of ICON (ICON-LAM). The preparation of ICON-CLM was triggered at DWD in the project ProWaS (Projection Service for Waterways and Shipping) – a joint pilot program of several German Federal Agencies – to prepare a regular federal forecasting and projection service about the influence of climate change on coastal and waterway traffic.

ICON can be used on a wide range of scales from climate projection, climate prediction, numerical weather prediction (NWP) down to large-eddy simulations (Heinze et al., 2017). For these different scales, there is a number of different modes as shown in Figure 1. Generally, there are three different physics packages available: the NWP, the ECHAM physics, and the large-eddy physics (~~which is not shown in Figure 1). The first one, named ICON-NWP,~~ LES physics). Within the first physics package, at global scale, the ICON-NWP is used for operational weather forecasting at DWD. ~~The second one, which is called~~ ICON-EU-Nest is the regional ICON on the European domain nested within the global ICON-NWP. ICON-LAM denotes the limited-area mode of ICON-NWP, which currently is available for the NWP and large-eddy configurations. ICON-LEM (ICON-Large Eddy Model) applies the physics package dedicated to large eddy simulations to study processes such as cloud, convection and turbulence on a very high resolution grid. Within the ECHAM physics package, the global atmospheric model ICON-A (Giorgetta et al., 2018), originating from the general circulation model ECHAM6~~atmospheric model~~, is used for global climate simulations. This configuration is coupled to the global ocean model ICON-O (Korn, 2017) and the land and biosphere model JSBACH (Brovkin et al., 2013) within the ICON Earth System Model (ICON-ESM). ~~A~~ The feature for one- or two-way nested sub-domains with grid refinement ~~is~~ available in the NWP ~~configuration and~~ physics package (ICON-EU-Nest) has recently also been transferred to ~~ICON-A~~ the ECHAM package by DWD (ICON-EUClim). ~~ICON-LAM denotes the limited-area mode of ICON, which currently is available for the NWP and large-eddy configurations.~~

ICON-CLM builds upon ICON-LAM, and currently contains a set of technical adaptations for climate applications.

The aim of this paper is to introduce the new regional climate mode of ICON~~(~~, ICON-CLM~~)~~, along with its starter package ICON-CLM_SP, a supporting infrastructure needed to perform long-term simulations. ~~Tests with different model time steps~~ Different technical tests and tests on the impact of prescribing upper boundary conditions interpolated from re-analysis data were carried out. A long evaluation simulation driven by ERA-Interim ~~re-anlysis~~ re-analysis (Dee et al., 2011) was conducted over a period of ~~30~~ 20 years and the results were compared to the evaluation simulation of the latest ~~recommended~~ COSMO-CLM version ~~(~~recommended by the CLM-Community (called "recommended version", CCLM 5.0 clm9). The paper is structured as follows: Some related general information on ICON-NWP and ICON-LAM is provided in Section 2. The adaptations in model source code and technical infrastructure are described in Section 3. Section 4 gives details of the ICON-CLM model configuration and setup for the evaluation run as well as the evaluation methods we used. Results of ~~this~~ the technical tests and of the evaluation run in comparison to observational data and to the results of the latest COSMO-CLM version are shown in Section 5. Conclusions are provided in Section 6.

2 General information on ICON-NWP and ICON-LAM

All ICON models in the ICON family (Figure 1) utilize unstructured triangular grids that originate from a spherical icosahedron with 20 equal sized triangles. ICON horizontal grid is denoted as R_nB_k; this is a triangular grid generated from the original icosahedron by first dividing the edges into n parts, followed by k subsequent edge bisections. The division of the edges into n equal parts gives n^2 spherical triangles within the original triangle. In the second step, each triangle is again subdivided into 4 smaller triangles. The resulted grid R_nB_k has, therefore, the total number of triangle cells n_{cells} and the number of edges n_{edges} calculated from the following formulas:

$$n_{cells} = 20n^2 4^k; \quad n_{edges} = 30n^2 4^k \tag{1}$$

The effective grid size $\overline{\Delta x}$ is defined as:

$$\overline{\Delta x} \approx 5050/(n2^k) \quad [km] \tag{2}$$

Some characteristics of the model grids used and mentioned in this work are listed in Table 1. Figure 2c visualizes the R2B8 grid extracted from the EURO-CORDEX domain (marked in red box from Figure 2a) .

The vertical layer distribution in the ICON models is a height-based coordinate system following the terrain, with denser layers near the earth surface and gradually changing to constant height model levels above a certain height. Two options for the height-based terrain-following vertical coordinate are offered in the ICON models, the terrain-following hybrid Gal-Chen coordinate (Simmons and Burridge, 1981) and the Smooth Level Vertical SLEVE coordinate (Schär et al., 2002; Leuenberger et al., 2010). With SLEVE (used in the simulations in this paper), the influence of small-scale terrain features decays more quickly with height than the large-scale features in order to obtain smooth vertical coordinate levels at mid and upper levels. The vertical coordinate is a function of model top height, the layer thickness of the lowermost layer, the total number of vertical layers, and the stretch factor which controls the distribution of the model levels. Users can define the vertical model levels by setting these controlling parameters via the model namelist.

Outputs in the ICON models can be written out in GRIB or Net-CDF format. Options for outputting on the ICON native grid or regular lat-lon grid or rotated lat-lon grid are available. Outputs can be written with individual or multiple fields in an output file. Users can define how many output steps in one output file and the output frequency. It is possible to have outputs with different intervals, for example hourly precipitation, daily temperature and monthly mean mean sea level pressure (MSLP) in the same run.

At the lateral boundary of the limited area domain, a sponge layer is applied, within which the internal flow is gradually relaxed towards the external boundary data. At the outer most area of the limited domain, the "lateral boundary zone" is a stripe fixed with 4 cell rows. Here the external boundary data are simply prescribed. After the outer 4 cell rows is the "lateral boundary nudging zone". The width of this nudging zone can be defined in the namelist setting of ICON with the minimum value of 8 cell rows to prevent the boundary artifacts. The nudging coefficient gradually reduces from the outer to the inner edge of the "lateral boundary nudging zone", making the influence of the prescribed external data weaker. The strength of the nudging can be controlled by the maximum relaxation coefficient in the model namelist.

For the upper boundary, ICON-LAM offers an option of prescribing the upper boundary conditions by using the same driving data source as for the lateral boundary conditions (nudging option). Users can define the height of the nudging zone as well as the nudging coefficients for the horizontal wind and for the thermodynamic variables via namelist settings. If this vertical nudging option is turned off, a Rayleigh damping is applied to the vertical wind speed within the damping layer in order to prevent unphysical reflection of vertically propagating gravity waves.

3 Model development

As the Limited-Area Mode of ICON, which ICON-CLM builds upon, has originally been developed for NWP applications, several adaptations and technical extensions were necessary to prepare the model for climate applications. Apart from the adjustments in the code, long-term climate simulations require a technical infrastructure for data and job management. Such an infrastructure has also been developed based on the existing infrastructure of COSMO-CLM.

3.1 The regional climate model ICON-CLM

Weather forecasting, which predicts the state of the atmosphere only up to about 2 weeks in advance, often does not involve the development of the ocean state. The ocean surface condition, hence, is often kept constant during the forecasts in weather prediction models or just slightly adjusted with a climatological trend for the forecast period. Thus, in ICON-LAM there is only option to update the sea surface temperature (SST) and sea-ice cover ~~in ICON-LAM can only be updated~~ monthly. For

ICON-CLM, ~~it is necessary~~ we want to have a ~~more frequent (up to hourly)~~ flexible option to update of SST and sea ice from external data at a user-defined interval. For this purpose, an option for ~~higher~~ flexible update frequencies of these boundary conditions was implemented in ICON-CLM. Time-dependent SST and sea-ice data can now be read from external data files and are fed to ICON-CLM with ~~an user-defined interval~~ shorter intervals than one month (e.g. 1 hourly or 6 hourly). The user can select this option of frequent update of SST and sea ice via namelist settings. The external SST and sea ice data must be prepared and remapped to the ICON grid.

Similarly, the ~~green house~~ greenhouse gas (GHG) values are usually kept constant in weather forecast models, because the changes during the forecast period are negligible. In climate projections, however, it is necessary to use the time-dependent GHGs ~~from climate change~~ as provided by corresponding GHG scenarios. Such an option was already available in ICON, but only in combination with the ECHAM physics package. The corresponding read routine was therefore extended so that it works for the NWP physics as well. Some additions to the NWP radiation scheme were made with respect to the GHG vertical profile with a new option to ~~get~~ retrieve the profile from external gas data. A file that contains yearly values of CO_2, CH_4, N_2O and Chlorofluorocarbons (CFC) for all years of the experiment needs to be prepared in advance. These features of the time-dependent SST and GHG were largely based on the corresponding implementations in the ICON-A (Giorgetta et al., 2018).

~~For the upper boundary, ICON-LAM offers an option of prescribing the upper boundary conditions by using the same driving data source as for the lateral boundary conditions (nudging option). Users can define the height of the nudging zone as well as the nudging coefficients for the horizontal wind and for the thermodynamic variables via namelist settings. If this vertical nudging option is turned off, a Rayleigh damping is applied to the vertical wind speed within the damping layer in order to prevent unphysical reflection of vertically propagating gravity waves.~~

In the NWP configuration of ICON-LAM, the number of soil layers is always constant with eight layers. The depths of half soil layers are also fixed at values between 5 mm and 14.5 m. However, for climate simulations in domains other than Europe (e.g. Africa, Asia) or to achieve better simulation of the soil variables for the European domain, it is usually reasonable to adjust these soil parameters. Therefore, an option for a flexible number and depth of the soil layers has been implemented in the ICON-CLM code.

The input/output of ICON-CLM has also been adjusted to have more flexibility. In NWP mode, the precipitation data are accumulated from the start of the forecast till the end without any reset. This is suitable for short weather forecasts, but for long climate simulations, this procedure is inconvenient and could, in the worst case, cause problems due to data imprecision. Furthermore, the maximum and minimum 2-m temperature values are calculated for 6 hourly intervals in NWP applications, while for climate simulations the standard for these output variables is usually 24 hours. To control this flexibility ~~extensions~~extension, new namelist parameters were introduced in ICON-CLM.

At the lateral boundaries, ICON-LAM requires, by default, information on cloud liquid water content and cloud ice water content from the global forcing data. These input fields are usually available if the ICON-CLM lateral boundary conditions are taken from reanalysis data like ERA-Interim. But if global climate projections are used as lateral boundary conditions, these fields are usually not provided. Thus, the model code has been adjusted so that if cloud liquid water content and cloud ice content are not available in the lateral boundary data, these variables are initialized with zero.

Table 2 provides an overview of some differences between COSMO-CLM and ICON-CLM.

3.2 The starter package ICON-CLM_SP

In order to facilitate long-term climate simulations, we developed a run time infrastructure called starter package and a separate evaluation tool. Both are provided along with the ICON-CLM model source code. The starter package ICON-CLM_SP contains a run routine, a climatological testsuite, all necessary utilities and configure scripts for different super computing environments. At the moment, two system settings for ~~Cray~~ Nec-Aurora (DWD) and Atos/Bull (DKRZ) are supported and tested (note that our ICON-CLM simulations in this paper were done on the DWD Cray XC40, this computer was replaced by the Nec-Aurora afterwards). Settings for other machines could be easily added if necessary.

The run routine in ICON-CLM_SP, called "subchain", was adapted from the routine of the existing COSMO-CLM package. The "subchain" contains five sub-routines for input preparation (prep), converting input data (conv2icon), ICON-CLM job management (icon), archiving (arch) and postprocessing (post) of the model output. Sub-routine "prep" copies and checks all the global forcing data as input for "conv2icon". Then "conv2icon" preprocesses and interpolates the initial data and the lateral, lower and upper boundary data onto the ICON-CLM model grid for the current model simulation. Sub-routine "icon" does the job management for ICON-CLM model. After that, all model output data are compressed by "arch" and some post-processing steps like the provision of time series of selected output variables are done in "post". Usually the simulations in ICON-CLM are done per calendar month with restarts at the end of each month, however the simulation duration can be easily customized by changing in the sub-routine "icon".

A climatological ~~testsuite~~ test suite (CTS) was also created based on the CTS from COSMO-CLM. In the CTS, 5-year test simulations can be done automatically with "subchain". The users can choose one simulation as a reference. The test simulations then will be compared with the reference simulation with respect to observational data (E-OBS and CRU, see Section 4.2 for more details) by an extra sub-routine called "eval". At the end, the results are visualized with standardized plots. This CTS was built for the purpose of testing different versions of model source codes, or different setups of the same model version. Hence, it is a very helpful tool for model development and tuning.

Besides the sub-routine "eval" in CTS, a separate evaluation tool called "ETOOLS" was also adapted from the COSMO-CLM evaluation tool. This tool provides comparisons of the simulation results with observation data sets and creates standardized plots to visualize the results. In order to facilitate the transition from COSMO-CLM to ICON-CLM for the users, both ICON-CLM_SP and ETOOLS were created such that the “look and feel” as well as the usage of the software packages is as similar as possible to the corresponding packages that exist for the COSMO-CLM model. The output structure of ICON-CLM or post-processed time series from "subchain/post" are also similar to those of COSMO-CLM for the same reason. ~~Furthermore~~On this account, users should be able to use all existing scripts and programs that were developed for COSMO-CLM output also for ICON-CLM data.

4 Data and methods

4.1 Model configuration and experiment setup

~~After the implementation of the necessary changes in the model source code, a number of technical tests was performed. First, the influence of different domain decompositions for parallelization on the results were tested at the super computer at DWD (Cray XC40). Results were binary identical, independent of the domain decomposition and the number of processors used for the simulations. Repeating tests starting from the same restart state were also carried out. ICON-CLM also showed binary identical results here.~~

All of the ICON-CLM tests described afterwards and the evaluation run were performed at the resolution R2B8 (approximately 10 km) on a domain (Figure 2a) completely covering the ~~EU-CORDEX~~ EURO-CORDEX domain (Giorgi et al., 2009)~~(Figure 2, left)~~. The model atmosphere is divided into 70 vertical layers and the model top is at a height of 30 km or 10 hPa. The soil in ICON-CLM contains eight layers down to a depth of 14.5 m. ICON-CLM was driven at the lateral and lower boundaries by ERA-Interim at 6 hourly intervals. The model atmosphere was initialized with ERA-Interim data. The soil temperature and soil moisture were taken from a previous ~~test~~ simulation which is long enough to ensure that the spin-up of the soil has been completed. The monthly Tegen aerosol climatology (Tegen et al., 1997) and the ozone climatology from Global Earth system Monitoring using Satellite and in-situ data (GEMS) were used in our simulations. For the upper boundary, as described in Section ~~3~~2, there are two options: (1) Using the driving data and nudging gradually in the relaxation zone; (2) Damping the vertical wind beneath the upper boundary. To assess the impact of these two options, two 10-year simulations (1979-1988) were done with the same ~~setups~~set-up, with and without global data nudging. Analysis from these 10-year runs resulted in very minor differences on surface variables and none of the options showed any advantage over the other. For the evaluation run, we chose the option with nudging data at the upper boundary with ERA-Interim data, as later we wanted to compare the results with those from a COSMO-CLM run using a similar nudging option. The nudging zone started from the height of 12 km to the model top of atmosphere (30 km). Variables which are nudged within this layer is the horizontal wind and the thermodynamic variables (air pressure and temperature) with nudging coefficient 0.04 and 0.075 respectively.

In order to find a suitable model configuration for ICON-CLM at the resolution R2B8, an optimized namelist setup was used, namely the setup from ~~ICON-LAM~~ ICON-NWP for R3B7 with nested domain on R3B8 grid (approximately 13 km and 6.5 km respectively). The tuning parameters were taken over from the global settings. In this setting, the ~~Tiedke~~Tiedtke/Bechtold (Bechtold et al., 2008) convection parameterization scheme and the Rapid Radiation Transfer Model (RRTM) radiation scheme (Mlawer et al., 1997) were used. These setups were checked to make sure that they are appropriate for climate applications and were used in all simulations.

Former simulations (not published) with COSMO-CLM showed that in some cases the model results ~~depend~~ depended on the chosen model time steps. There ~~is~~ was one particular time step that ~~leads~~ led to larger biases in precipitation and surface pressure, especially over the Alps and the south western area of the ~~model~~ EURO-CORDEX domain. This issue has been ~~analyzed~~ analysed by the CLM-Community but is still not fully understood yet. To ensure that such a dependency of the results on the model time step is not present in ICON-CLM, different fast physics/advection time step (hereafter: time step) choices were tested. At R2B8 (approximately 10 km) resolution, the time step should not exceed 120 seconds for stability reasons. With the common model and experiment setups described above, we carried out multiple one-year simulations for the year 1979 with time steps of 60, 80, 90, 100 and 120 seconds. Figure 3 shows the biases compared to reference data of 2-m temperature, ~~mean sea level pressure (MSLP), total precipitation~~total precipitation, MSLP and total cloud cover. The biases were averaged for each month and for the Alpine region (sub-region denoted AL in Figure 2~~, right~~c). Colors show the different time step experiments. The biases for all variables from any particular time-step experiment were small and did not stand out from the rest. ~~The results were similar for all~~ Similar results were found for all other sub-regions ~~shown in Figure 2 (right)~~ and therefore are not shown here. The annual and seasonal biases of these multiple one-year simulations were also very similar ~~in all cases~~ with all time steps (not shown here).

Because there is no big difference in the model results depending on the choice of time step, for experiments at spatial resolution R2B8, we chose the time step of 90 seconds due to the computational efficiency and stability, the time step of COSMO-CLM at similar horizontal resolution is 100 seconds. An evaluation run was carried out for the ~~EU-CORDEX~~ EURO-CORDEX domain at the resolution R2B8. The simulation period is 1979 to ~~2016.~~ 2000. The model and experiment setups were the common setups as described above, this evaluation run is later referred to as ICLM-REF.

The results of ICLM-REF were compared to the reference experiment of the recent recommended version of COSMO-CLM (v5.0_clm9). This COSMO-CLM simulation is later referred to as CCLM-REF. This COSMO-CLM setup showed the best performance for the ~~European~~ EURO-CORDEX domain in an inter-comparison with ~~several other~~ a large number of setups which was performed within the COPAT project (COordinated Parameter Tuning~~) project~~ , a project providing a thorough evaluation of a large number of COSMO-CLM configurations to come up with a recommended version) in the CLM-Community. The simulation period is 1979 to ~~2000.~~ 2000; the model resolution is 0.165°, also about 10 km like in ICLM-REF. The initial, lower and lateral boundary data are taken from ERA-Interim. A sponge layer with Rayleigh damping in the upper levels of COSMO-CLM domain was used. The damping was done against the external boundary values, similarly to the nudging at the model top in ICLM-REF simulation and thus the results from both experiments are comparable.

4.2 Evaluation methods

For model assessment and evaluation, output fields from six variables were ~~analyzed~~analysed, namely 2-m temperature, daily maximum and minimum values of 2-m temperature, MSLP, total precipitation and total cloud cover. Monthly average values of these variables were calculated and used for further analysis. For total precipitation, the monthly accumulated amounts were calculated. Parts of the evaluation were carried out based on seasonal averages. The following definitions and abbreviations of the seasons are used in this paper: winter - December, January, February (DJF), spring - March, April, May (MAM), summer - June, July, August (JJA), and autumn - September, October, November (SON). Results were averaged for eights sub-regions as already used in the PRUDENCE projects (described by Christensen and Christensen (2007)). These sub-regions are: ~~Britian Iles~~ British Isles (BI), Iberian Peninsula (IP), France (FR), Mid-Europe (ME), Scandinavia (SC), Alps (AL), Mediterranean (MD) and Eastern Europe (EA) (shown ~~on Figure 2, right~~in Figure 2c).

For the evaluation of ~~monthly~~ mean values of 2-m temperature, daily maximum and minimum 2-m temperature, MSLP and total precipitation, the E-OBS dataset (Haylock et al., 2008; Van den Besselaar et al., 2011) was used as reference data. E-OBS is a 0.25° gridded daily dataset covering all of Europe. The data are available over land ~~and available~~ from quite recent back to 1950. In order to compare 2-m temperature data from different datasets (~~ICON-CLM, COSMO-CLM~~ ICLM-REF, CCLM-REF and E-OBS), a height correction was performed. The model temperature values at E-OBS 2-m height were calculated based on the differences between model and E-OBS surface elevation and the moist adiabatic lapse rate (0.0065 K/m). The reference cloud data, which was used for assessment of the model cloud cover, are CRU TS data (Harris et al., 2014). This is a monthly gridded dataset at 0.5° resolution, available globally over land area. The dataset covers the period from 1901 to 2013. ~~Because the outputs of ICLM-REF and CCLM-REF were written on rotated grids that are~~

For all ICON-CLM simulations in this paper, the outputs were written in Net-CDF format and on the rotated lat-lon grid as in CCLM-REF. Because this rotated lat-lon grid is finer than E-OBS and CRU grids, these data were ~~transformed to the~~ remapped to the regular lat-lon grids ~~of the observational data or~~ with the same spatial resolution as the observational data for the purpose of comparison. The E-OBS and CRU datasets contain data only over land, therefore the evaluation in this paper (e.g. areal averaged fields) were done using only land points.

The evaluation within COPAT for 2-m temperature~~and MSLP~~ , MSLP and cloud cover was done using E-OBS version 10.0 and CRU version 3.22 as reference data. Therefore, in order to compare our evaluation to the one from COPAT, we also used the same versions of the data sets. The comparison period is 20 years from 1981 to 2000, the same as the evaluation period in COPAT. The reference total precipitation data was taken from E-OBS version 12.0 because this dataset (among versions from 10.0 to 17.0) shows the fewest missing data for precipitation over the area of Poland.

Some important climate indices (listed in Table 3) were calculated from ~~ICON-CLM, COSMO-CLM~~ ICLM-REF, CCLM-REF and E-OBS 2-m temperature and total precipitation data ~~for the entire~~ and averaged over the period 1981-2000. The number of days that fulfils the definition (in Table 3) was counted for each horizontal grid cell, then averaged over a sub-region.

Root-mean-square error (RMSE) was calculated from the model and observed monthly values (monthly aggregated values for precipitation and monthly mean values for other variables):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(S_i - O_i)^2} \tag{3}$$

where S_i and O_i are the model (ICLM-REF or CCLM-REF) and reference data (E-OBS or CRU TS) monthly values, respectively, averaged over the sub-regions considered at the ith month; N is the total number of months in the evaluation period 1981-2000.

To compare the spatial variability of the data, ~~spatio-temporal~~ spatial standard deviation (STDEV) was also calculated from ~~monthly~~ time-averaged fields of model and observed data~~for all simulation months, and then averaged over time.~~. The STDEV ratio ($STDEV_{model}/STDEV_{observation}$) was calculated for the ease of comparing the spatial variation of the two model data with respect to the observation. The model with STDEV ratio closer to one better represents the spatial variation of the observation data.

5 ~~Evaluation and comparison with COSMO-CLM~~

~~In the evaluation run,~~ The results of air temperature from ICLM-REF ~~showed a very good performance . In comparison with reference data ,~~ were also compared to those of other regional climate models (RCMs) within the EURO-CORDEX experiments. A thorough evaluation of several EURO-CORDEX ensembles is presented in Kotlarski et al. (2014). One ensemble, called EUR-11, has quite similar set-up to our ICLM-REF ~~biases are already of similar magnitude as the CCLM-REF biases. This result is consistent for all six evaluated variables~~ set-up. The horizontal resolution of the RCMs in this EUR-11 ensemble is about 12 km; these simulations were also driven by ERA-Interim. The simulation period, however, differs from ours, 1989-2009 instead of 1981-2000. Nevertheless, this comparison for sure gives us some knowledge about ICON-CLM performance relatively to other state-of-the-art RCMs. In Figure 5 and ~~for all PRUDENCE sub-regions. For certain areas or certain variables~~ B1 in Kotlarski et al. (2014)), some similar analysis to ours were done for 9 different RCMs, seasonal means of temperature bias are shown for 8 PRUDENCE sub-regions. These biases were also calculated against E-OBS data like in our evaluation, though an older version of E-OBS data were used in their paper. These figures were used to compare to our figures from ICLM-REF.

5 Testing and evaluating ICON-CLM

5.1 Technical tests

After the technical adaptation in the ICON model source code to enable long-term climate simulations, a number of technical tests was performed. First, the influence of different domain decompositions for parallelization on the simulation results was tested at the super computer Cray XC40 at DWD. The domain of ICON models is split by a built-in geometric subdivision according to the number of processors used for calculation. Tests with ICON-CLM were done using 120, ~~ICLM-REF showed even smaller biases than~~ 240, 480, 960 computer processors, each test simulated the year 1979. As these tests are one year long each, they contain multiple monthly restarts of the model. The model outputs were checked for all climate variables evaluated in this paper. Results were binary identical, independent of the numbers of processors used for the simulating. ICON-CLM showed the ability to produce the same results and to restart properly with a variety of domain decompositions.

Repeating tests starting from the same restart state were also carried out. ICON-CLM was restarted at the time point 01.01.1980 at 00:00 UTC multiple times from the same restart file on the Cray XC40. Each simulation was carried out for one month from 01.01.1980 to 31.01.1980. Model outputs were checked for all model calculation steps and for all climate variables evaluated in this paper. In addition, the restart files created at the end of each test for the simulation time 01.02.1980 at 00:00 UTC were also compared. The results from these repeating tests showed binary identical values.

Two additional restart tests were also done. In the first test, ICON-CLM was run for two months from 01.01.1980 to 29.02.1980 without restarting in between. And in the second test, ICON-CLM also simulated these two months but was restarted at 01.02.1980 00:00 UTC. The model outputs were compared, with focus on the period after the restart of the second test. Results showed that restarting did not introduce any difference in model diagnostics.

Short tests up to few months were also done on a different computing system, the Atos/Bull at the German Climate Computing Centre (DKRZ). ICON-CLM showed the ability to run stably on at least two machines. Performance tests were done for ICON-CLM and COSMO-CLM on the Atos/Bull. The two models were run in one-month simulations on European domain. The horizontal resolution is roughly 50 km, same number of computer processors was used. ICON-CLM was about 15% faster than COSMO-CLM in these tests.

5.2 Evaluation and comparison with COSMO-CLM

Initially a longer simulation with ICON-CLM (30 years) was planned and carried out. But since the data from CCLM-REF ~~.~~

5.3 ~~Air temperature~~

are only available to us in the period 1981-2000, only data in this period were taken for evaluation.

5.2.1 Air temperature

The 2-m temperature bias of ICLM-REF and CCLM-REF was mostly within -1.5 K to 1.5 K relative to E-OBS data. Figure 4 shows the mean annual biases against E-OBS data over the 20 year period from 1981 to 2000 for ICLM-REF and CCLM-REF data for the entire domain. ~~Biases over central Europe, especially Southern Germany and the Alps, from both experiments were of the same sign, with smaller biases coming from~~ Both experiments had dominant negative bias over Eastern Europe, with ICLM-REF to a lesser extend. ICLM-REF had a warm bias in most parts of Sweden and southern Russia, while CCLM-REF showed a cold bias in these regions. The Balkan region is well known for large air temperature biases in COSMO-CLM simulations (Anders and Rockel, 2009; Pham et al., 2014; Trusilova et al., 2014); which also ~~occur~~ occurred in ICLM-REF with values up to ~~+~~1.5 K.

The seasonal temperature bias in Figure 5 shows that the median of the bias in the sub-regions ranged from ~~-0.6~~-0.58 K to ~~0.8~~0.81 K for ICLM-REF and ~~-1.1~~-1.13 K to ~~1.2~~1.20 K for CCLM-REF. While in many of the sub-regions and seasons ~~nearly no~~ relatively small bias was found for ICLM-REF (bias median in the order of 0.01 K), CCLM-REF had usually larger biases. Both models had smaller biases in winter and autumn and larger biases in spring and summer. Besides some sub-regions like ~~Britain~~British Isles, Mid-Europe, France, Iberian Peninsula where small biases were found (especially from ICLM-REF), sub-regions Scandinavia, Alps, Mediterranean and Eastern Europe showed larger biases. The latter sub-regions also showed larger variability of biases in space, with the bias range from -3.7 K to 3.4 K. CCLM-REF had almost always more variability in temperature bias than ICLM-REF. The differences in spatial variability in CCLM-REF were exceptionally strong in Scandinavia in winter and summer.

~~For extreme daily~~ Compared with the nine RCMs in EURO-CORDEX ensemble EUR-11 (Kotlarski et al., 2014), ICLM-REF showed similar magnitude of biases for all four seasons. Our ICLM-REF results from Figure 5 were placed in comparison to Figure 5 and B1 in Kotlarski et al. (2014). ICLM-REF and other EURO-CORDEX experiments tended to have negative biases for air temperature in winter. In spring the opposite was observed, ICLM-REF gave positive bias for most of the sub-regions, while EUR-11 continued to have negative biases. In all seasons, ICLM-REF biases stayed well fit within the spread of the EUR-11 ensemble, and had smaller biases compared with some other RCMs. In winter, for example, it can be seen on the spread of the solid circles (representing EUR-11 ensemble) in Figure 5 and B1 in Kotlarski et al. (2014) that the temperature biases of EUR-11 ensemble were up to -4 K and -3 K for sub-region Alps and Mediterranean respectively; and for other sub-regions up to around -2 K. Meanwhile, ICLM-REF had mean biases quite closed to 0 for most of the sub-regions (Figure 5, DJF).

For daily max/min temperature bias, CCLM-REF and ICLM-REF showed opposite ~~trends~~results. CCLM-REF tended to underestimate the maximum 2-m temperature, while ICLM-REF overestimated the values. This can ~~most clearly be~~ be most clearly seen in winter and autumn (Figure 6), where the medians of almost all ICLM-REF biases ~~are~~ were positive and of ~~all~~ most CCLM-REF biases ~~are~~ were negative. ICLM-REF clearly had smaller median biases ranging from -0.02 to ~~1 K(with one exception of 2.3 K)~~2.34 K, while the range in CCLM-REF was -2.5 to 1.2 K. In summer, the bias was larger for both models. CCLM-REF had larger spatial differences among the sub-regions than ICLM-REF. Similarly, ~~ICLM-REF simulated the~~ the bias of the minimum 2-m temperature ~~better than~~ of ICLM-REF was reduced compared to CCLM-REF ~~with median biases ranging from~~ (-1.2 K to 0.5 K compared to -0.6 K and 1.8 ~~Kin CCLM-REF (~~K, Figure 7). ICLM-REF slightly underestimated the daily minimum 2-m temperature while CCLM-REF overestimated it. Consequently, the daily temperature range was overestimated by ICLM-REF and underestimated by CCLM-REF. However, the representation of the daily temperature range is, in general, in ICLM-REF closer to the observed values than in CCLM-REF. The overestimation of the diurnal range of temperature near the surface in ICLM-REF was probably caused by a positive radiation bias, which is known in ICON. This positive bias is especially larger when the radiation scheme RRTM (Rapid Radiative Transfer Model) (Mlawer et al., 1997; Barker et al., 2003) is used, which was the case in ICLM-REF (see Table 2). Recently, another radiation scheme, named ecRad (Rieger et al., 2019), is added into ICON. The positive bias for radiation is strongly reduced with the use of this radiation scheme.

ICLM-REF had also smaller RMSE than CCLM-REF for 2-m temperature over most sub-regions with one exception in sub-region Iberian Peninsula (Table 4). RMSEs of ICLM-REF ~~were all smaller than 0.9~~did not exceed 0.87 K, whereas CCLM-REF had RMSE up to 1.03 K. Looking at the STDEV ~~,~~ ratio, ICLM-REF had stronger spatial variation than E-OBS in three out of eight sub-regions and weaker variation in four out of eight. In some sub-regions, the STDEV ratio was equal or very close to one (France, Iberian Peninsula, Mediterranean). Compared with CCLM-REF~~tended to overestimate the spatial variability. This can be most clearly seen in sub-region Eastern Europe with STDEV of 9.03~~, ~~8.46 and 8.22 from CCLM-REF,~~ ICLM-REF ~~and E-OBS, respectively. ICLM-REF STDEV were closer to the E-OBS values than CCLM-REF STDEV in all~~ showed in six sub-regions ~~. The differences were especially pronounced in sub-regions Iberian Peninsula and France.~~ better STDEV ratios.

Similar results were found for minimum 2-m temperature, with ~~ICLM-REF´s RMSEs~~ RMSE from ICLM-REF between 0.36 and ~~STDEVs closer to the values of E-OBS~~ 0.85 K. All ICLM-REF's RMSEs were smaller compared to those from CCLM-REF, especially for sub-region Mediterranean and British Isles ICLM-REF errors were less than half (Table 5). ICLM-REF's STDEV ratio was closer to 1 for all sub-regions ~~(Table 5) .~~ in comparison with CCLM-REF. Both models did not simulate well the minimum 2-m temperature spatial variation in mountainous area (Scandinavia and Alps) as for flatter area, but did not show a tendency for specific type of orography.

Regarding RMSEs of maximum 2-m temperature (Table 6), ~~in~~ ICLM-REF showed larger biases compared to min/mean 2-m temperature with a maximum bias of 1.54 K in the Mediterranean sub-region. In four out of eight sub-regions, ICLM-REF got lower errors and ~~seven~~ five sub-regions showed better spatial variation. Overall, ICLM-REF simulated average and daily ~~extreme~~ max/min values of 2m air temperature better than CCLM-REF.

The ~~better~~ improved representation of daily ~~extreme temperatures~~ max/min 2-m temperature in ICLM-REF resulted in a ~~better performance~~ reduced bias for climate indices ~~that~~ which are determined by air temperature. Among those indices, CCLM-REF overestimated the ~~total~~ annual numbers of ice days and tropical nights ~~over the whole~~ averaged over the evaluation period (1981-2000) as can be seen in Figure 11. Largest tropical night overestimation was found for the sub-region Eastern Europe with ~~104 nights~~ 5.2 nights per year by CCLM-REF, 6.5 times more than that of E-OBS (~~16 nights~~ 0.8 nights per year), while the ICLM-REF result was much closer to the observations with only ~~23 nights~~ 1.15 nights per year. Beside that, ICLM-REF ~~results~~ tropical nights indices were very close to the observed numbers for sub-regions France, Alps, Mediterranean, while the numbers in CCLM-REF clearly stand out against the observations. The results for the annual ice days index ~~is similar.~~ were similar; CCLM-REF overestimated the number of ice days for all sub-regions. ICLM-REF, on the other hand, slightly underestimated the annual numbers of ice days but was in all ~~regions~~ sub-regions much closer to the number derived from E-OBS than CCLM-REF. ~~The representation of frost days was much better in CCLM-REF compared to the previous two indices, but~~ Generally, ICLM-REF ~~still showed a better performance than~~ showed an underestimation of the annual frost days compared to observations over Europe, except for the Scandinavia sub-region. Compared to CCLM-REF~~in all eight sub-regions. Both models produced fewer frost days than observed, except for~~ , however, the underestimation was reduced in ICLM-REF. The biggest improvement was simulated in the sub-region ~~Scandinavia.~~ Alps.

The number of annual summer days was overestimated by ICLM-REF for six ~~of~~ from the eight sub-regions, while CCLM-REF mostly underestimated the amount of summer days. The strongest overestimation of the summer day index compared to CCLM-REF and E-OBS was seen in the Mediterranean sub-region. For three out of eight sub-regions and on average over the whole Europe, ICLM-REF simulated summer day indices more in agreement with observations compared to CCLM-REF. On average, for whole Europe, ICLM-REF resulted in ~~1065 summer days~~ 53.25 summer days per year; the numbers from E-OBS and CCLM-REF are ~~1128 and 874~~56.4 and 43.7, respectively.

5.3 ~~Precipitation~~

5.2.1 Precipitation

The mean annual precipitation bias ranged mostly from -50 mm/month to 50 mm/month in both models (Figure 4). Overall, both models simulated more precipitation than E-OBS data. However, one should keep in mind that E-OBS precipitation data tend to ~~give~~ suffer from gauge undercatch and evaporation leading to too low values (Gampe and Ludwig, 2017). According to Kotlarski et al. (2014), the systematic undercatch in E-OBS precipitation data can be on averaged in the order of 4-50 %. That might be the reason why both ICLM-REF and CCLM-REF overestimated precipitation in large part of the domain. CCLM-REF tended to produce too little precipitation than E-OBS along the Atlantic coast while ICLM-REF ~~performed better~~ agreed better with observation in this area. In the western part of Germany, for example, ICLM-REF had only a slight bias ~~with a difference of less than~~ of 5 mm/month compared with the reference data. CCLM-REF, on the other hand, produced negative biases up to -20 mm/month in this area. However, over the eastern part of Germany, ICLM-REF had larger biases up to more than 10 mm/month. In all other regions, the annual spatial distribution of precipitation biases was quite similar in both models.

Looking at the spatial variability of the seasonal biases ~~within~~ among the sub-regions in Figure 8, we see that although for some sub-regions in certain seasons, the bias medians were close to zero, the ranges of biases were large. This is expected because precipitation is a highly inhomogeneous variable. Summer and autumn tended to have small median bias in ICLM-REF and CCLM-REF. Among the four seasons, summer had the smallest variation probably due to the low precipitation amount. For winter, summer and autumn, it is not clear which model performed better. For spring, the median and the range of the bias were ~~better~~ in better agreement with observations in CCLM-REF compared to ICLM-REF for most of the sub-regions.

RMSE for precipitation from ICLM-REF ranged from 10.48 to 30.52 mm (Table 7). The largest error appeared over the Alps sub-region, probably due to the complicated terrain and the dependency of precipitation on orography. RMSEs of CCLM-REF were smaller than those of ICLM-REF for most of the sub-regions, except for sub-region Iberian Peninsula ~~(Table 7).~~ and Scandinavia. ICLM-REF simulated larger variation of precipitation in space compared to E-OBS data, with most STDEV ratio larger than 1, except for British Isles with 0.8. The spatial variability of precipitation in CCLM-REF was ~~also~~ closer to observations ~~with better STDEV for five~~ in six out of eight sub-regions indicated by an STDEV closer to one.

ICLM-REF tended to have more precipitation days than CCLM-REF, with the ~~wet days~~ annual wet day index higher for most of the sub-regions and only one exception for sub-region Scandinavia (Figure 11). In six sub-regions, ICLM-REF was more in line with observation than CCLM-REF. ICLM-REF also produced more days with heavy and very heavy precipitation on yearly average than CCLM-REF. For most of the sub-regions, both models overestimated heavy and very heavy precipitation indices, but CCLM-REF was often closer to E-OBS. From our results, ~~it is difficult to judge which model~~ CCLM-REF performed better for precipitation and precipitation related climate indices. ~~The results depended on the area~~ ICLM-REF had in part improvement in certain area of the model domain as well as ~~the season and the index(heavy/very heavy precipitation) considered~~for certain season and climate index.

5.3 ~~MSLP and cloud cover~~

5.2.1 Mean sea level pressure

The bias of MSLP of the two models showed opposite signs. ICLM-REF had positive biases while the biases in CCLM-REF were mostly negative as revealed in Figure 4. ~~Although both models had absolute~~ ICLM-REF had biases up to ~~2.5~~6 hPa, ~~the~~ while CCLM-REF had up to - 4 hPa. The performance of CCLM-REF was better~~for the sub-regions .~~, especially over Scandinavia and Iberian Peninsula. The opposite signs of the bias in the two model experiments can also be seen very clearly ~~on~~ in Figure 9. Spring and autumn showed less spatial variability of MSLP values in both models. The representation of MSLP was in winter and autumn ~~slightly~~ better in CCLM-REF than in ICLM-REF with smaller median bias and smaller bias ranges.

The RMSEs and ~~STDEVs~~ STDEV ratios for MSLP are shown in Table 8. ICLM-REF had better RMSE than CCLM-REF in half of the sub-regions. However, ICLM-REF had some large errors over sub-regions Iberian Peninsula and France with RMSEs of ~~1.9~~1.90 hPa and ~~1.7~~1.71 hPa, respectively. ~~The performance of both models was very similar regarding the spatial variation of MSLP~~Regarding spatial variation, ICLM-REF gave better STDEV ratio than CCLM-REF in most of the sub-regions. Both model resulted in largest difference in STDEV with respect to E-OBS data over sub-region Alps with the ratio of 1.76 for ICLM-REF and 2.09 for CCLM-REF.

Additionally, we evaluated the representation of the large-scale MSLP in ICON-CLM by looking at its spatial MSLP pattern. The driving data ERA-Interim was used for reference and ICLM-REF result was also compared to CCLM-REF. The mean spatial fields of MSLP from the three data sets were calculated for the evaluation period 1981-2000. All data were regridded to ERA-Interim grid to ensure a fare comparision and are shown in Figure 12, the illustration areas are different among the data sets due to the different model domains. As can be seen, ICLM-REF could be able to reproduce the large-scale pattern of MSLP from the forcing data. The pressure structure looks quite similar between ICLM-REF and ERA-Interim, with the low and high pressure systems located in the right areas. Because of the small domain of ICLM-REF compared with ERA-Interim, it is hard to view the whole high pressure system located west of Portugal. But from what we can see, the highest pressure was a bit closer to the coast line in ICLM-REF than in ERA-Interim. This unfortunately was out of CCLM-REF domain. ICLM-REF underestimated the MSLP near Iceland by several hectopascal. CCLM-REF reproduced the MSLP better than ICLM-REF in this area with the magnitude and pattern very similar to those from ERA-Interim.

5.2.2 Cloud cover

The representation of annual mean cloud cover in ICLM-REF looks much better than in CCLM-REF in Figure 4. CCLM-REF produced too much cloud ~~cover over the sub-region~~ over Scandinavia and over the eastern part of the domain (up to 0.2, or in percentage 20%, more cloud cover than CRU TS data). ICLM-REF, on the other hand, had biases of only up to +/- 10%. For most regions, the bias was in the range of +/- 5%. The overestimation of cloud cover in the sub-regions Scandinavia and Eastern Europe might be the reason for the cold bias of CCLM-REF in these regions ~~. On a seasonal mean~~ (see Figure 4).

On seasonal mean as shown in Figure 10, however, CCLM-REF showed a smaller negative bias during winter in all sub-regions (except Scandinavia), and its performance in autumn was comparable to ICLM-REF. ~~The RMSEs and STDEVs~~ ICON-CLM simulated cloud cover noticeably better than COSMO-CLM over the sub-region Scandinavia.

The RMSEs of cloud cover did not show much difference between the two ~~models. No concrete conclusion could be drawn from those numbers , therefore they were not shown here.~~ simulations as presented in Table 9. Same conclusion can be drawn for STDEV ratios, numbers were somewhat similar from both ICLM-REF and CCLM-REF.

5.3 Large-scale information reproduced by ICON-CLM

As the limited area models are only forced by the global data at their boundaries, inside the limited domain, the regional climate models can more or less freely develop their own circulation. One of the major concerns in regional climate modelling is to what extend the regional climate models modify the large-scale information from the global data (Sanchez-Gomez et al., 2009).

We want to test if ICON-CLM can be able to reproduce the large-scale atmospheric condition from the forcing data ERA-Interim. The geopotential at 500 hPa data from ICLM-REF and ERA-Interim were averaged at different temporal scales, 6 hourly, daily, monthly, seasonal and yearly. Due to the different grids and ERA-Interim lower resolution, ICLM-REF data were remapped to ERA-Interim grid. The spatial correlation coefficient was calculated for each time slot within the evaluation period 1981-2000.

Time series of the correlation coefficient for the different temporal scales are shown in Figure 13. For all of the time scales we considered, results showed higher correlation for longer time scale, the correlation is lowest for 6 hourly data and strongest for yearly data. For the 6 hourly and daily means, the mean correlation coefficient over the whole time period 1981-2000 was high, 0.925 and 0.928 respectively. But the correlation for some time slots dropped below 0.8, causing the minimum values of correlation coefficient only over 0.5. ICON-CLM showed correlation coefficients above 0.85 for longer time scales from monthly to yearly. With this result, we can conclude that ICON-CLM reproduced well the large scale of the driving data for time scales from monthly to yearly, and partially for time scales from 6 hourly to daily.

6 Conclusions

The new regional climate model ICON-CLM has been derived from the weather forecast model ICON-LAM along with the necessary technical infrastructure and evaluation tools allowing users to carry out and evaluate long-term regional climate simulations. An evaluation run from the very first version of ICON-CLM showed very promising results. The ICON-CLM results were proven to be independent of the domain decomposition, and restarting with the same configuration gave binary identical results. ~~In contrast to some versions of to COSMO-CLM,~~ ICON-CLM did not show any systematic dependency of the results on integration time steps which was found in COSMO-CLM. All tested time steps showed similar results with no bias outliers for any of the chosen values. These time step tests were done with model horizontal resolution R2B8 (about 10 km) over the ~~EU-CORDEX~~ EURO-CORDEX domain. When choosing another model grid spacing or simulating another domain, tests might be required to affirm these results.

The vertical nudging of the global forcing data at the model upper boundary did not show any notable impact on the climatology of surface variables. This is probably due to the fact that in our settings of ICON-CLM, the top of the atmosphere was at 30 km height. When choosing a lower model top, one might see larger effects of the vertical nudging. Furthermore, in this study, we focused on the near-surface climate and therefore did not look at the upper air layers where larger differences between nudging and no nudging ~~are~~ can be found.

Results from the evaluation run showed that ICON-CLM ~~performed already as well as COSMO-CLM. Especially for air temperature and~~ was able to reproduce the large scale atmospheric circulation from the driving model and the most important climate variables. In comparison with reference data, ICLM-REF biases are of similar magnitude, for certain areas or certain variables, even slightly smaller than CCLM-REF biases. Improvements were visible in ICLM-REF for air temperature, and hence temperature-related climate indices~~, improvements are visible in the ICLM-REF simulation~~. The reason might be that ~~ICON-CLM simulated better cloud cover with less overestimation~~ ICLM-REF showed a lower overestimation of cloud coverage over the northern and eastern part of the domain ~~than COSMO-CLM~~compared to CCLM-REF. For precipitation and MSLP, the ~~performance of both models was very similar.~~ results from CCLM-REF were in better agreement with E-OBS data.

It should be taken into account that ICON-CLM is still in the early stage of development in climate mode while COSMO-CLM has been developed and applied for NWP and climate applications for more than 20 years and was well-tuned with a large number of tested namelist combinations in the COPAT project. Therefore, ICON-CLM has still great potential to improve ~~the~~ its model setup and with growing experience we expect that ICON-CLM results will improve further in the upcoming years.

Comparison for air temperature between ICLM-REF and the EURO-CORDEX ensemble showed that ICLM-REF biases landed in the upper part of the ensemble bias spreads. A broader inter-comparison with EURO-CORDEX ensemble is recommended once an optimum set-up for ICON-CLM is established.

As written above, only short tests were done on a different computing system other than the Cray XC 40 at DWD. More technical tests, therefore, should be done on the Atos/Bull at DKRZ. Also a comparison of the results from different machines will show how ICON-CLM is dependent on the computer systems.

The next step in the ICON-CLM preparation will be a thorough model tuning by testing the sensitivity of the model to a variety of namelist parameters and ~~their~~ different combinations of namelist settings in order to find an optimal configuration. Climate simulations on different domains, e.g. CORDEX Africa, will be done to evaluate the ability of ICON-CLM to simulate different climates. So far only re-analysis driven ~~simulation~~ simulations have been performed with ICON-CLM, but historical simulations driven by the results of global climate simulations will also be done to test the model performance for this experiment type. Based on a well-evaluated model configuration, regional climate projections will be ~~performed~~ carried out in order to address the impact of climate change at regional scale. Thus, climate projections, e.g. in the framework of CORDEX, will also be provided with ICON-CLM in the future. We plan to further develop ICON with the aim of unifying the different physics packages currently existing for the numerical weather prediction and the climate mode in order to pursue a "Seamless prediction" system with one forecasting system which can produce ~~produce~~ forecasts and projections for all time-scales from weather prediction to seasonal~~and~~ , decadal predictions and climate projections.

Code availability. To institutions, the ICON model is distributed under an institutional license issued by DWD. In case of the institutional license, two copies of the institutional license need to be signed and returned to the DWD. The ICON Release Version 2.6.1 can be then downloaded at https://data.dwd.de.

To individuals, the ICON Model is distributed under a personal non-commercial research license distributed by MPI-M. Every person receiving a copy of the ICON framework code accepts the ICON personal-non-commercial research license by doing so. Or, as the license states: Any use of the ICON-Software is conditional upon and therefore leads to an implied acceptance of the terms of the Software License Agreement. To receive an individually licensed copy, please follow the instructions provided at https://code.mpimet.mpg.de/projects/iconpublic/wiki/Instructions_to_obtain_the_ICON_model_code_with_a_personal_non-commercial_research_license. The ICON Release Version 2.6.1 model code with a personal non-commercial research license can be obtained at https://code.mpimet.mpg.de/projects/icon-downloads/files.

The starter package ICON-CLM_SP_beta1 can be downloaded from http://doi.org/10.5281/zenodo.3896136. The starter package will be shipped with a recommended configuration for EURO-CORDEX domain at the resolution R2B8 and will be available for all CLM-Community members.

The forcing data were used for our simulations are the ERA-Interim data (Dee et al., 2011). Model evaluation was done againt E-OBS version 10.0 and 12.0 (Haylock et al., 2008; Van den Besselaar et al., 2011) und CRU TS version 3.22 datasets (Harris et al., 2014).

Appendix A

A1

Author contributions. Trang Van Pham implemented the adaptations in model code and ICON-CLM Starter Package, carried out the experiments and prepared the manuscripts with contributions from all co-authors. Christian Steger contributed in the organization and supervision of the work. Ingo Kirchner and Mariano Mertens are the source code administrators of ICON-CLM. Burkhardt Rockel contributed to the ICON-CLM Starter Package. Klaus Keuler, Burkhardt Rockel and Barbara Früh defined the research goals, aims and methodology. Daniel Rieger and Günther Zängl contributed in the development of ICON-NWP, the foundation for ICON-CLM.

Competing interests. The authors declare that they have no conflict of interest.

Disclaimer. TEXT

Acknowledgements. The authors thank to all ICON development teams at DWD and MPI-M for providing us the ICON model code which is a great basis for our development. Thanks to the colleagues who provided the COSMO-CLM evaluation from COPAT project. Trang Van Pham acknowledged the financial support of DWD via project ProWaS. We acknowledged the E-OBS dataset from the EU-FP6 project ENSEMBLES (http://ensembles-eu.metoffice.com) and the data providers in the ECA&D project (http://www.ecad.eu)"

References

Anders, I. and Rockel, B.: The influence of prescribed soil type distribution on the representation of present climate in a regional climate model, Climate Dynam., 33, 177–186, 2009.

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M., and Reinhardt, T.: Operational convective-scale numerical weather prediction with the COSMO model: description and sensitivities, Monthly Weather Review, 139, 3887–3905, 2011.

Barker, H. W., Stephens, G., Partain, P., Bergman, J., Bonnel, B., Campana, K., Clothiaux, E., Clough, S., Cusack, S., Delamere, J., et al.: Assessing 1D atmospheric solar radiative transfer models: Interpretation and handling of unresolved clouds, Journal of Climate, 16, 2676–2699, 2003.

Bechtold, P., Köhler, M., Jung, T., Doblas-Reyes, F., Leutbecher, M., Rodwell, M. J., Vitart, F., and Balsamo, G.: Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time-scales, Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 134, 1337–1351, 2008.

Brovkin, V., Boysen, L., Raddatz, T., Gayler, V., Loew, A., and Claussen, M.: Evaluation of vegetation cover and land-surface albedo in MPI-ESM CMIP5 simulations, Journal of Advances in Modeling Earth Systems, 5, 48–57, 2013.

Christensen, J. H. and Christensen, O. B.: A summary of the PRUDENCE model projections of changes in European climate by the end of this century, Climatic change, 81, 7–30, 2007.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., et al.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc., 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Doms, G. and Schättler, U.: The nonhydrostatic limited-area model LM (Lokal-Modell) of DWD, Part I: Scientific documentation, Deutscher Wetterdienst (DWD), 10, 1999.

Doms, G., Förstner, J., Heise, E., Herzog, H., Mironov, D., Raschendorfer, M., Reinhardt, T., Ritter, B., Schrodin, R., Schulz, J.-P., et al.: A description of the nonhydrostatic regional COSMO model. Part II: Physical parameterization, Deutscher Wetterdienst, Offenbach, Germany, 2011.

Gal-Chen, T. and Somerville, R. C.: On the use of a coordinate transformation for the solution of the Navier-Stokes equations, Journal of Computational Physics, 17, 209–228, 1975.

Gampe, D. and Ludwig, R.: Evaluation of gridded precipitation data products for hydrological applications in complex topography, Hydrology, 4, 53, 2017.

Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., Brovkin, V., Crueger, T., Esch, M., Fieg, K., et al.: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5, Journal of Advances in Modeling Earth Systems, 5, 572–597, 2013.

Giorgetta, M. A., Brokopf, R., Crueger, T., Esch, M., Fiedler, S., Helmert, J., Hohenegger, C., Kornblueh, L., Köhler, M., Manzini, E., et al.: ICON-A, the atmosphere component of the ICON Earth System Model. Part I: Model Description, Journal of Advances in Modeling Earth Systems, 2018.

Giorgi, F., Jones, C., Asrar, G. R., et al.: Addressing climate information needs at the regional level: the CORDEX framework, World Meteorological Organization (WMO) Bulletin, 58, 175, 2009.

Harris, I., Jones, P. D., Osborn, T. J., and Lister, D. H.: Updated high-resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset, International journal of climatology, 34, 623–642, 2014.

Haylock, M., Hofstra, N., Klein Tank, A., Klok, E., Jones, P., and New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, J. Geophys. Res.: Atmospheres, 113, https://doi.org/10.1029/2008JD10201, 2008.

Heinze, R., Dipankar, A., Henken, C. C., Moseley, C., Sourdeval, O., Trömel, S., Xie, X., Adamidis, P., Ament, F., Baars, H., et al.: Large-eddy simulations over Germany using ICON: A comprehensive evaluation, Quarterly Journal of the Royal Meteorological Society, 143, 69–100, 2017.

Korn, P.: Formulation of an unstructured grid model for global ocean dynamics, Journal of Computational Physics, 339, 525–552, 2017.

Kotlarski, S., Keuler, K., Christensen, O. B., Colette, A., Déqué, M., Gobiet, A., Goergen, K., Jacob, D., Lüthi, D., Van Meijgaard, E., et al.: Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble, Geoscientific Model Development, 7, 1297–1333, 2014.

Leuenberger, D., Koller, M., Fuhrer, O., and Schär, C.: A generalization of the SLEVE vertical coordinate, Monthly Weather Review, 138, 3683–3689, 2010.

Majewski, D. and Ritter, B.: Das Global-Modell GME, Promet, 27, 111–122, 2002.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., and Clough, S. A.: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave, Journal of Geophysical Research: Atmospheres, 102, 16 663–16 682, 1997.

Orr, A., Bechtold, P., Scinocca, J., Ern, M., and Janiskova, M.: Improved middle atmosphere climate and forecasts in the ECMWF model through a nonorographic gravity wave drag parameterization, Journal of Climate, 23, 5905–5926, 2010.

Pham, T. V., Brauch, J., Dieterich, C., Früh, B., and Ahrens, B.: New coupled atmosphere-ocean-ice system COSMO-CLM/NEMO: assessing air temperature sensitivity over the North and Baltic Seas, Oceanologia, 56, 167–189, https://doi.org/10.5697/oc.56-2.167, 2014.

Rieger, D., Köhler, M., Hogan, R., Schäfer, S., Seifert, A., Lozar, A. d., and Zängl, G.: ecRad in ICON – Implementation Overview, Reports on ICON, 2019.

Ritter, B. and Geleyn, J.-F.: A comprehensive radiation scheme for numerical weather prediction models with potential applications in climate simulations, Monthly Weather Review, 120, 303–325, 1992.

Rockel, B., Will, A., and Hense, A.: The regional climate model COSMO-CLM (CCLM), Meteorol. Z., 17, 347–348, https://doi.org/10.1127/0941-2948/2008/0309, 2008.

Sanchez-Gomez, E., Somot, S., and Déqué, M.: Ability of an ensemble of regional climate models to reproduce weather regimes over Europe-Atlantic during the period 1961–2000, Climate Dynamics, 33, 723–736, 2009.

Schär, C., Leuenberger, D., Fuhrer, O., Lüthi, D., and Girard, C.: A new terrain-following vertical coordinate formulation for atmospheric prediction models, Monthly Weather Review, 130, 2459–2480, 2002.

Schrodin, R. and Heise, E.: The multi-layer version of the DWD soil model TERRA_LM, DWD, 2001.

Schulz, J.-P., Vogel, G., Becker, C., Kothe, S., Rummel, U., and Ahrens, B.: Evaluation of the ground heat flux simulated by a multi-layer land surface scheme using high-quality observations at grass land and bare soil, Meteorologische Zeitschrift, 25, 607–620, https://doi.org/10.1127/metz/2016/0537, http://dx.doi.org/10.1127/metz/2016/0537, 2016.

Seifert, A. and Beheng, K. D.: A double-moment parameterization for simulating autoconversion, accretion and selfcollection, Atmospheric research, 59, 265–281, 2001.

Simmons, A. J. and Burridge, D. M.: An energy and angular-momentum conserving vertical finite-difference scheme and hybrid vertical coordinates, Monthly Weather Review, 109, 758–766, 1981.

Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., Salzmann, M., Schmidt, H., Bader, J., Block, K., et al.: Atmospheric component of the MPI-M Earth System Model: ECHAM6, J. Adv. Model Earth Syst., 5, 146–172, https://doi.org/10.1002/jame.20015, 2013.

Tegen, I., Hollrig, P., Chin, M., Fung, I., Jacob, D., and Penner, J.: Contribution of different aerosol species to the global aerosol extinction optical thickness: Estimates from model results, Journal of Geophysical Research: Atmospheres, 102, 23 895–23 915, 1997.

Tiedtke, M.: A comprehensive mass flux scheme for cumulus parameterization in large-scale models, Monthly Weather Review, 117, 1779–1800, 1989.

Trusilova, K., Früh, B., Brienen, S., and Keuler, K.: Urban effects in climate simulations for CORDEX Europe, CLM-Community Newsletter, 2, 4–7, 2014.

Van den Besselaar, E., Haylock, M., Van der Schrier, G., and Klein Tank, A.: A European daily high-resolution observational gridded data set of sea level pressure, Journal of Geophysical Research: Atmospheres, 116, 2011.

Zängl, G., Reinert, D., Rípodas, P., and Baldauf, M.: The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core, Quarterly Journal of the Royal Meteorological Society, 141, 563–579, https://doi.org/https://doi.org/10.1002/qj.2378, 2015.

~~Overview ICON modelling framework.~~

~~(a) Simulation domain EURO-CORDEX and model orography mof ICLM-REF on the R2B8 grid. (b) Illustration of the icosahedral grid of ICON-CLM at the resolution R2B8 from a closer look at the red marked region from sub-figure (a). (c) Evaluation was done for the eight PRUDENCE sub-regions (BI: British Isles, IP: Iberian Peninsula, FR: France, ME: Mid-Europe, SC: Scandinavia, AL: Alps, MD: Mediterranean and EA: Eastern Europe).~~

~~Monthly averaged biases from ICON-CLM simulations with different time steps for 2-m temperature, total precipitation, MSLP compared to E-OBS data and for total cloud cover compared with CRU TS data. Data were averaged for the Alps (AL) sub-region and for year 1979.~~

~~Multi-year averaged biases over the period 1981-2000 against E-OBS data for 2-m temperature, total precipitation, MSLP and against CRU TS data for total cloud cover (from top to bottom respectively). Data from ICLM-REF and CCLM-REF evaluation runs (left and right respectively).~~

~~Seasonal mean 2-m temperature biases of ICLM-REF and CCLM-REF data against E-OBS data for all PRUDENCE sub-regions. Values averaged for the period 1981 to 2000. Spatial variability within a sub-region is indicated by the lower bar (5th), upper bar (95th), lower edge of the box (25th), middle of the box (50th) and upper edge of the box (75th percentile of the distribution of all grid points within a sub-region).~~

~~Same as Figure 5 but for daily maximum 2-m temperature.~~

~~Same as Figure 5 but for daily minimum 2-m temperature.~~

~~Same as Figure 5 but for total precipitation.~~

~~Same as Figure 5 but for MSLP.~~

~~Same as Figure 5 but for total cloud cover.~~

~~Area mean of climate indices days per yearcalculated from daily maximum and minimum air temperature and daily precipitation data of ICLM-REF, CCLM-REF and E-OBS for the PRUDENCE sub-regions and for the whole Europe (denoted EU) averaged over the period 1981-2000.~~

~~Mean sea level pressure (MSLP, in hPa) from ERA-Interim (left), ICLM-REF (middle) and CCLM-REF data (right). Data were average over the period 1981-2000.~~

~~Correlation coefficient of geopotential at 500 hPa between ICLM-REF and ERA-Interim data for various temporal scales, 6 hourly, daily, monthly, seasonal and yearly. Data were taken in the period 1981-2000.~~

~~Total numbers of triangle cells, edges and effective grid resolution of the grids mentioned in this paper. Grid Domain Number of triangle cells Number of triangle edges Effective grid resolution kmR2B8 EURO-CORDEX 286 824 430 988 9.9 R3B7 global 2 949 120 4 423 680 13.2 R3B8 Europe nested in global R3B7 659 156 989 911 6.6~~

~~Comparison between COSMO-CLM and ICON-CLM models. COSMO-CLM ICON-CLM Planetary boundary layer scheme Prognostic turbulent kinetic energy closure Wave dissipation at critical level Cumulus convection scheme Shallow convection: Reduced Tiedtke scheme for shallow convection only Deep convection: oder Tiedtke-Bechtold scheme from IFS Model Mass flux scheme with shallow, deep or mid-level convection. CAPE (convective available potential energy) based closure for deep convection. Boundary layer equilibrium closure for shallow convection. Large-scale omega (vertical velocity in~~

~~pressure coordinates) based closure for mid-level convection. Cloud microphysic scheme , reduced to one-moment scheme Single-moment scheme Radiation short and long wave δ two-stream radiation scheme after RRTM (Rapid Radiative Transfer Model) Land surface and soil scheme TERRA-ML Tiled TERRA Coordinate system horizontal: rotated geographical (lat/lon) vertical: terrain following Gal-Chen height coordinate and exponential height coordinate (SLEVE) according to horizontal: icosahedral grids vertical: terrain following Gal-Chen height coordinate and exponential height coordinate (SLEVE) according to~~

~~Description of climate indices. Index Description Unit Frost days index Number of days with minimum 2-m temperature $<$ $0°$C Days Ice days index Number of days with maximum 2-m temperature $<$ $0°$C Days Summer days index Number of days with maximum 2-m temperature $>$ $25°$C Days Tropical nights index Number of days with minimum 2-m temperature $>$ $20°$C Days Wet days index Number of days with total precipitation $\geq$ 1 mm Days Heavy precipitation days index Number of days with total precipitation $>$ 10 mm Days Very heavy precipitation days index Number of days with total precipitation $>$ 20 mm Days~~

~~RMSE and spatial STDEV ratio of averaged monthly 2-m temperature for the PRUDENCE sub-regions. Data from ICLM-REF, CCLM-REF and E-OBS from 1981 to 2000.~~

~~Same as Table 4 but for minimum 2-m temperature.~~

~~Same as Table 4 but for maximum 2-m temperature.~~

~~Same as Table 4 but for monthly accumulated total precipitation.~~

~~Same as Table 4 but for MSLP.~~

~~Same as Table 4 but for total cloud cover.~~

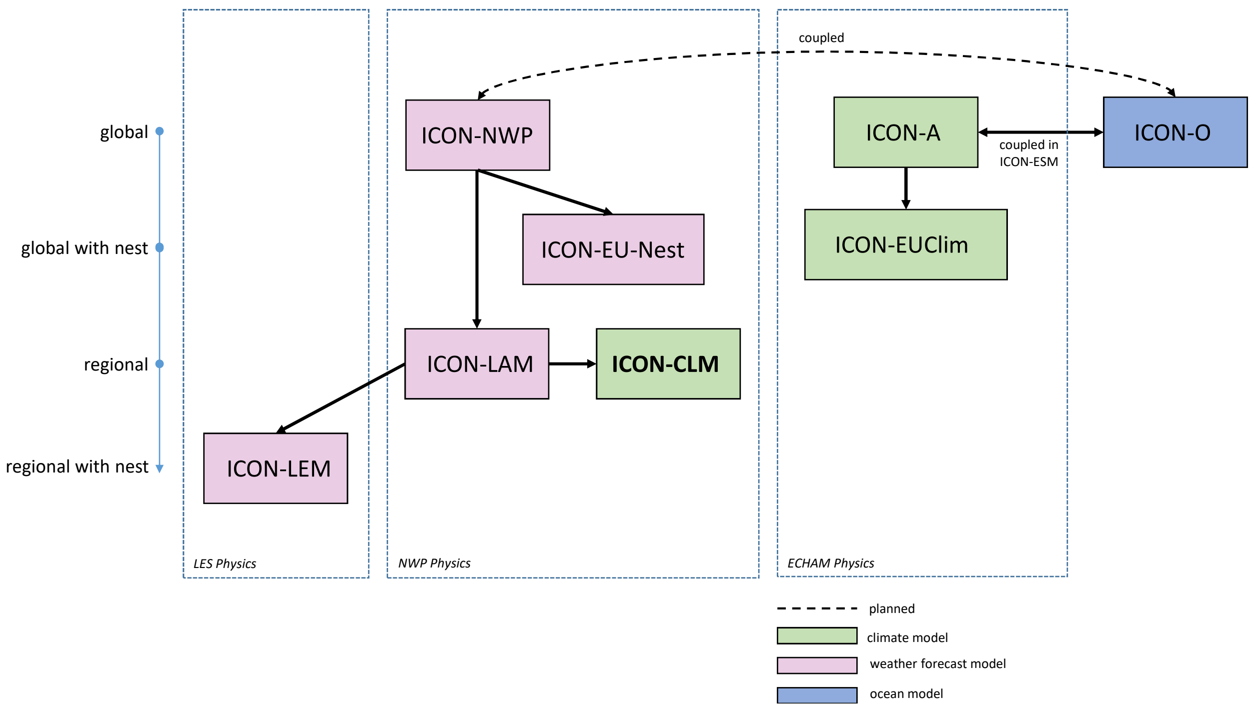

Figure 1. Overview ICON modelling framework.

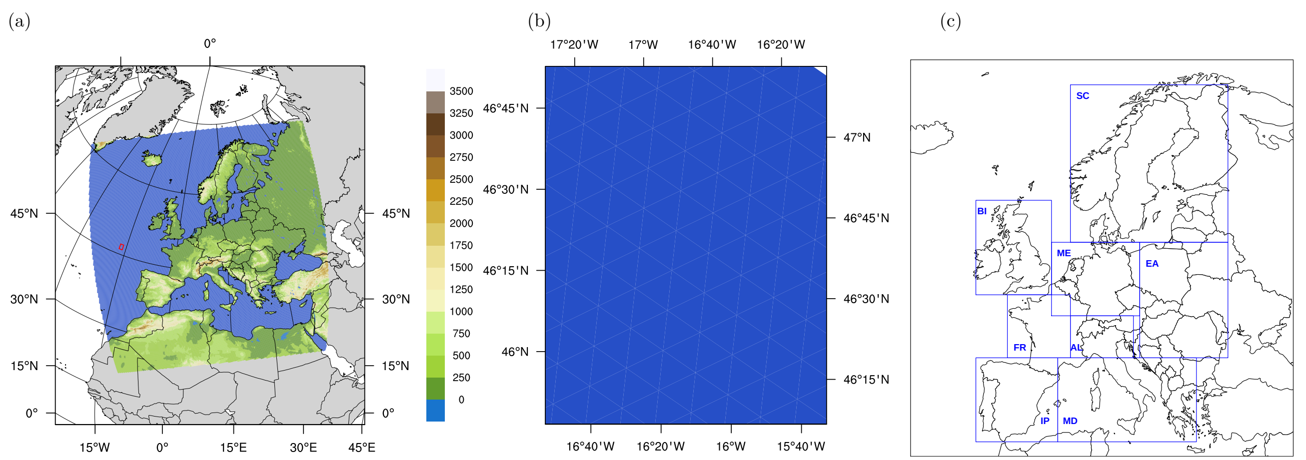

Figure 2. (a) Simulation domain EURO-CORDEX and model orography [m] of ICLM-REF on the R2B8 grid. (b) Illustration of the icosahedral grid of ICON-CLM at the resolution R2B8 from a closer look at the red marked region from sub-figure (a). (c) Evaluation was done for the eight PRUDENCE sub-regions (BI: British Isles, IP: Iberian Peninsula, FR: France, ME: Mid-Europe, SC: Scandinavia, AL: Alps, MD: Mediterranean and EA: Eastern Europe).

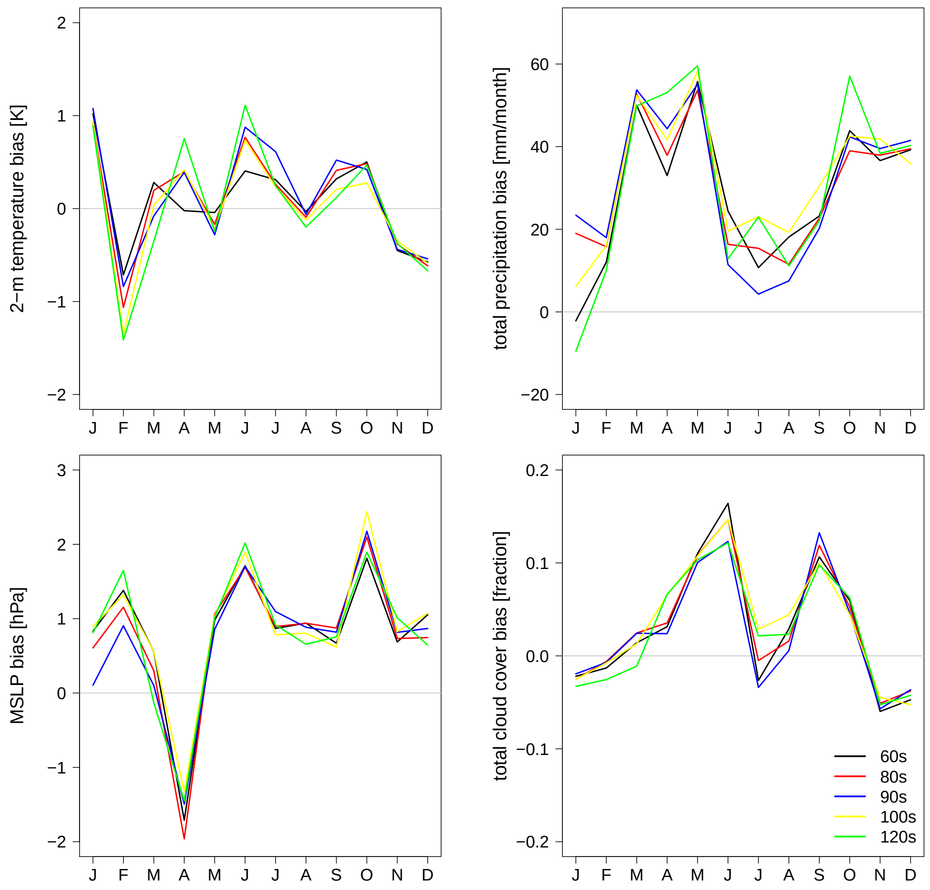

Figure 3. Monthly averaged biases from ICON-CLM simulations with different time steps for 2-m temperature, total precipitation, MSLP compared to E-OBS data and for total cloud cover compared with CRU TS data. Data were averaged for the Alps (AL) sub-region and for year 1979.

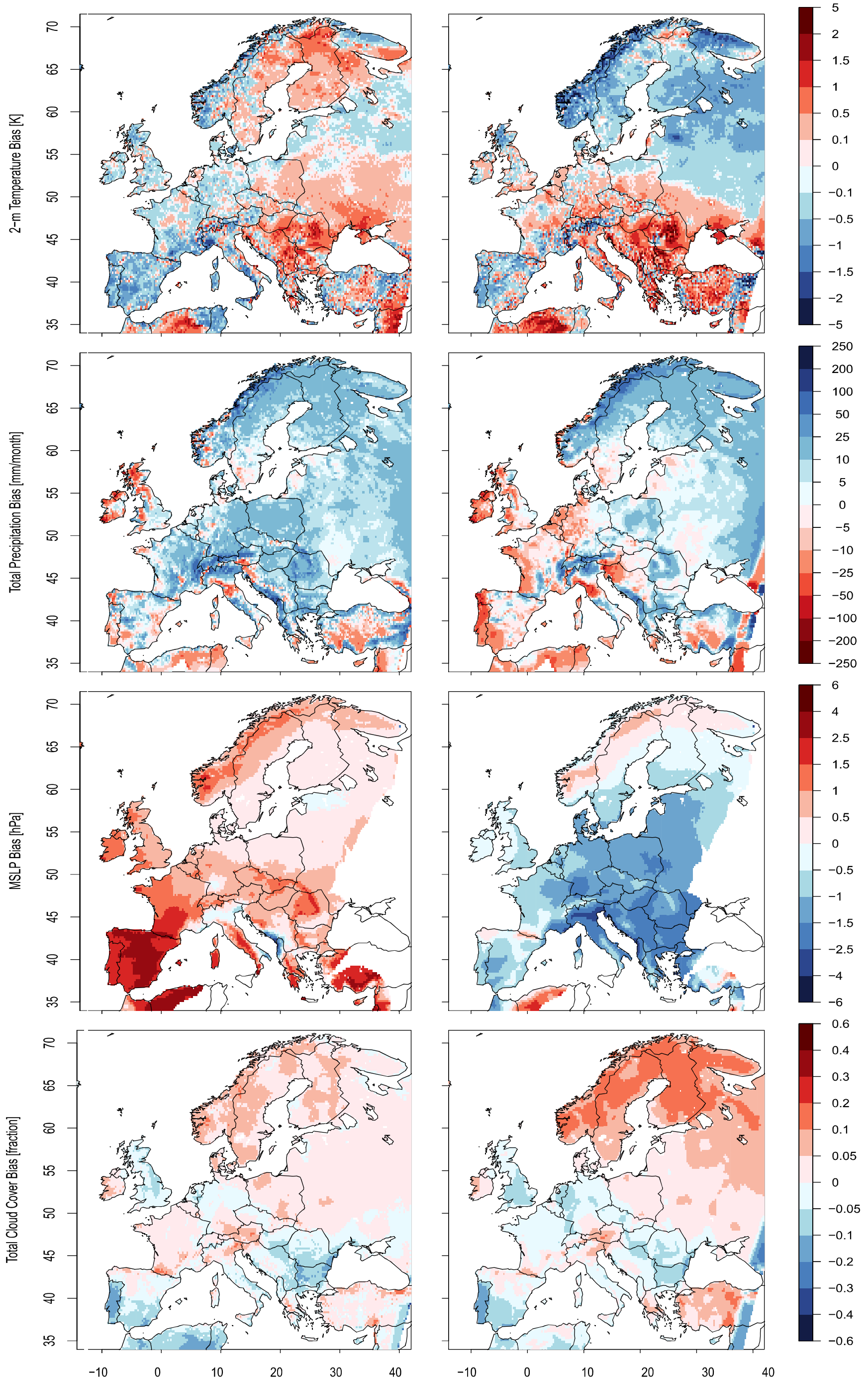

Figure 4. Multi-year averaged biases over the period 1981-2000 against E-OBS data for 2-m temperature, total precipitation, MSLP and against CRU TS data for total cloud cover (from top to bottom respectively). Data from ICLM-REF and CCLM-REF evaluation runs (left and right respectively).

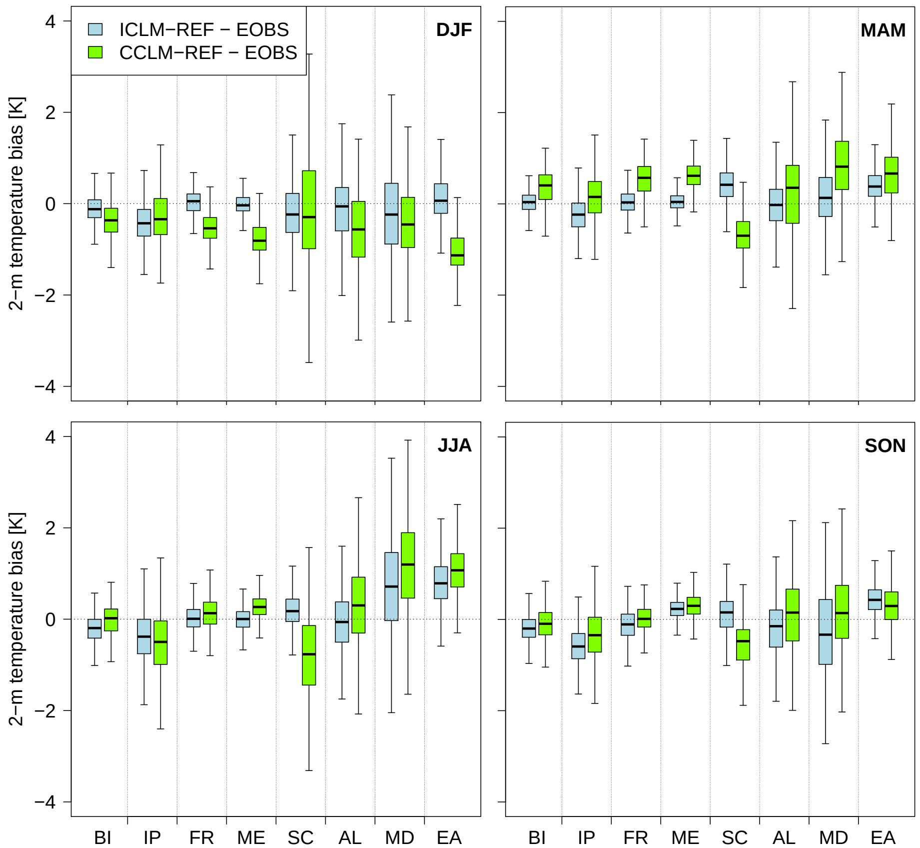

Figure 5. Seasonal mean 2-m temperature biases of ICLM-REF and CCLM-REF data against E-OBS data for all PRUDENCE sub-regions. Values averaged for the period 1981 to 2000. Spatial variability within a sub-region is indicated by the lower bar (5th), upper bar (95th), lower edge of the box (25th), middle of the box (50th) and upper edge of the box (75th percentile of the distribution of all grid points within a sub-region).

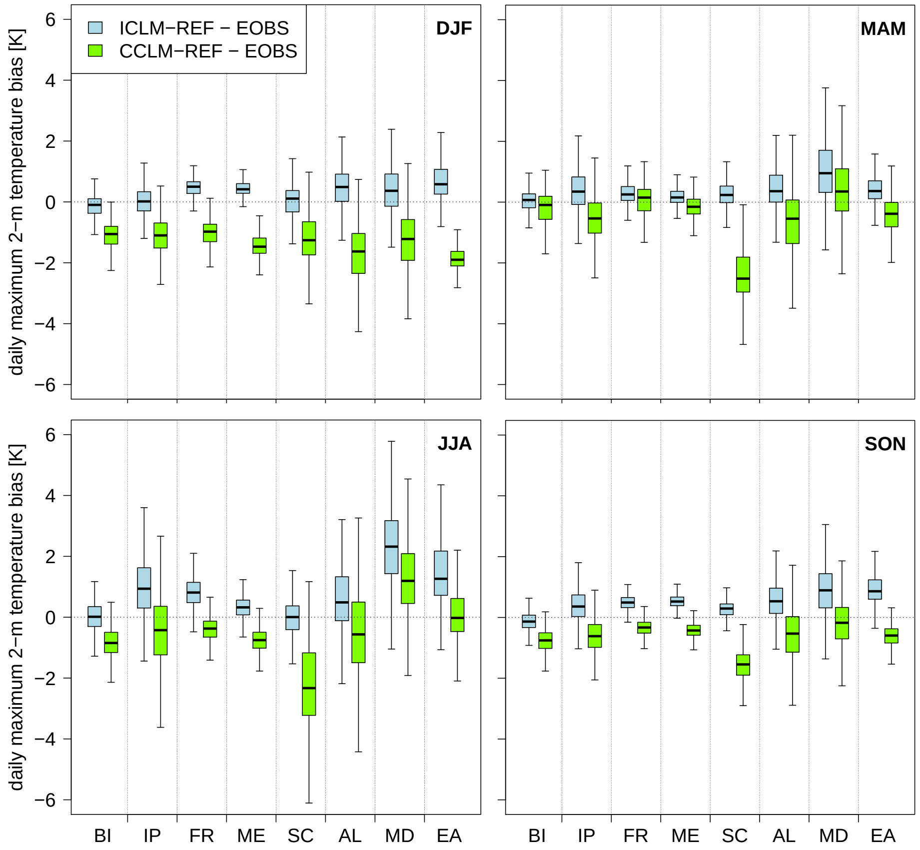

Figure 6. Same as Figure 5 but for daily maximum 2-m temperature.

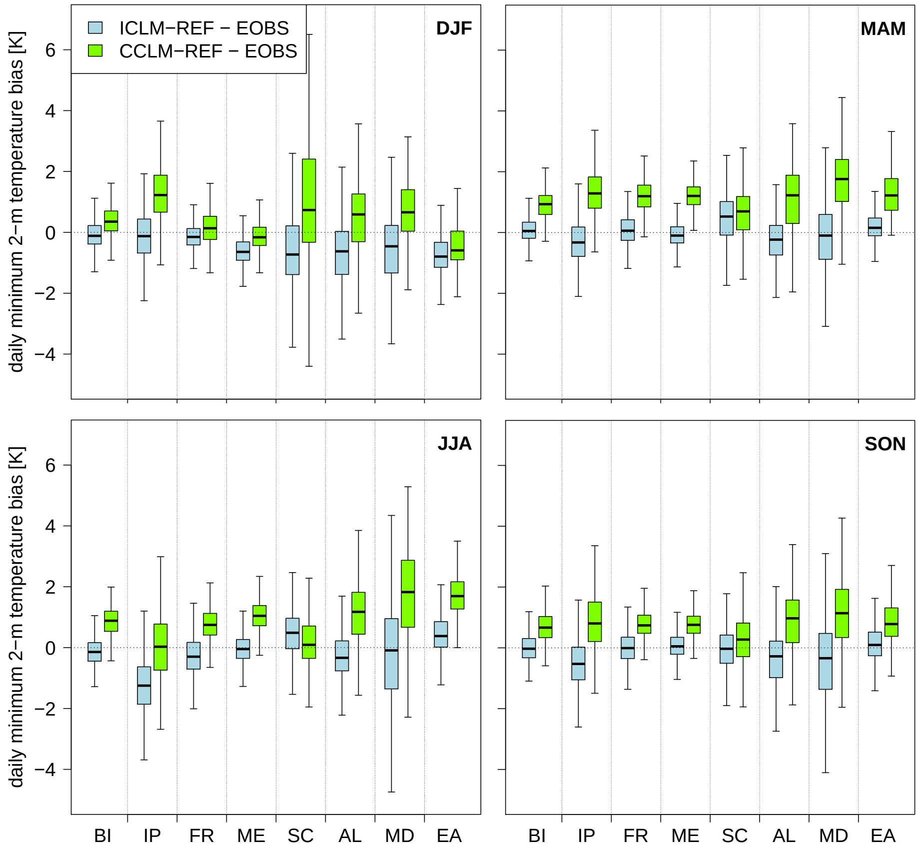

Figure 7. Same as Figure 5 but for daily minimum 2-m temperature.

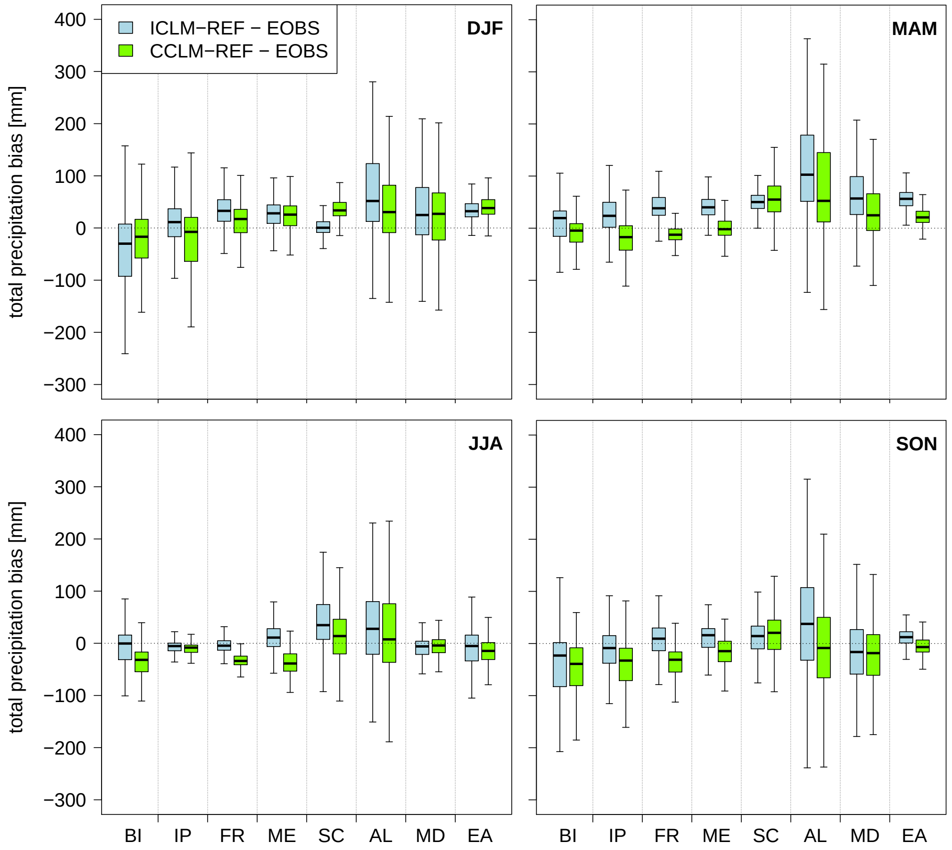

Figure 8. Same as Figure 5 but for total precipitation.

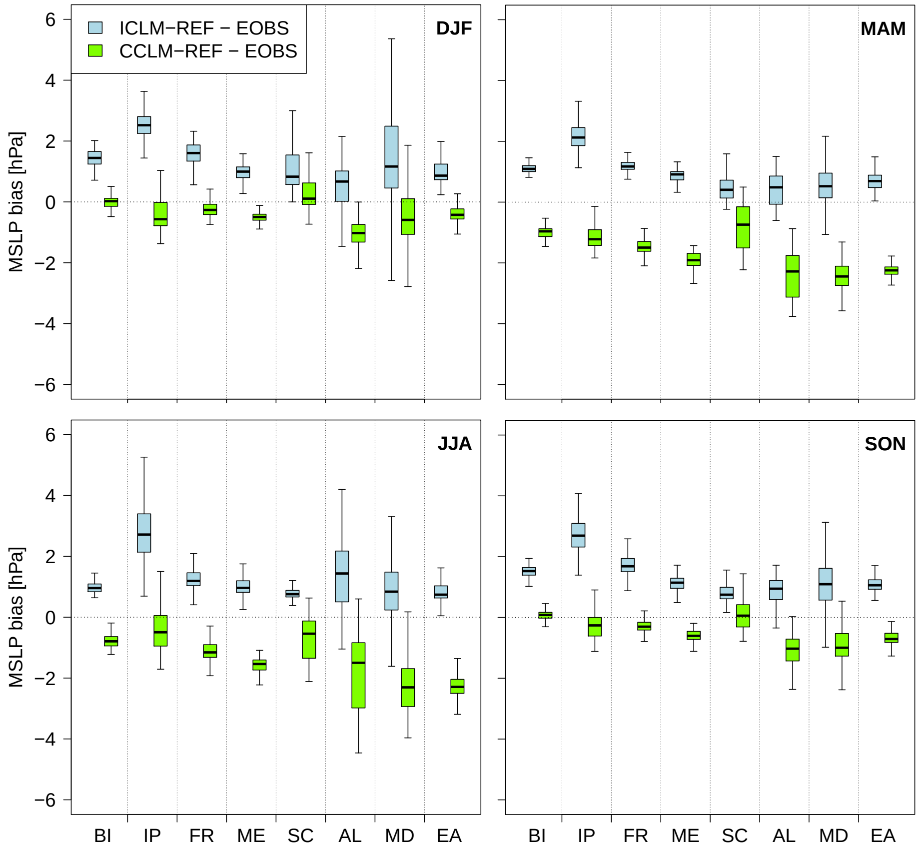

Figure 9. Same as Figure 5 but for MSLP.

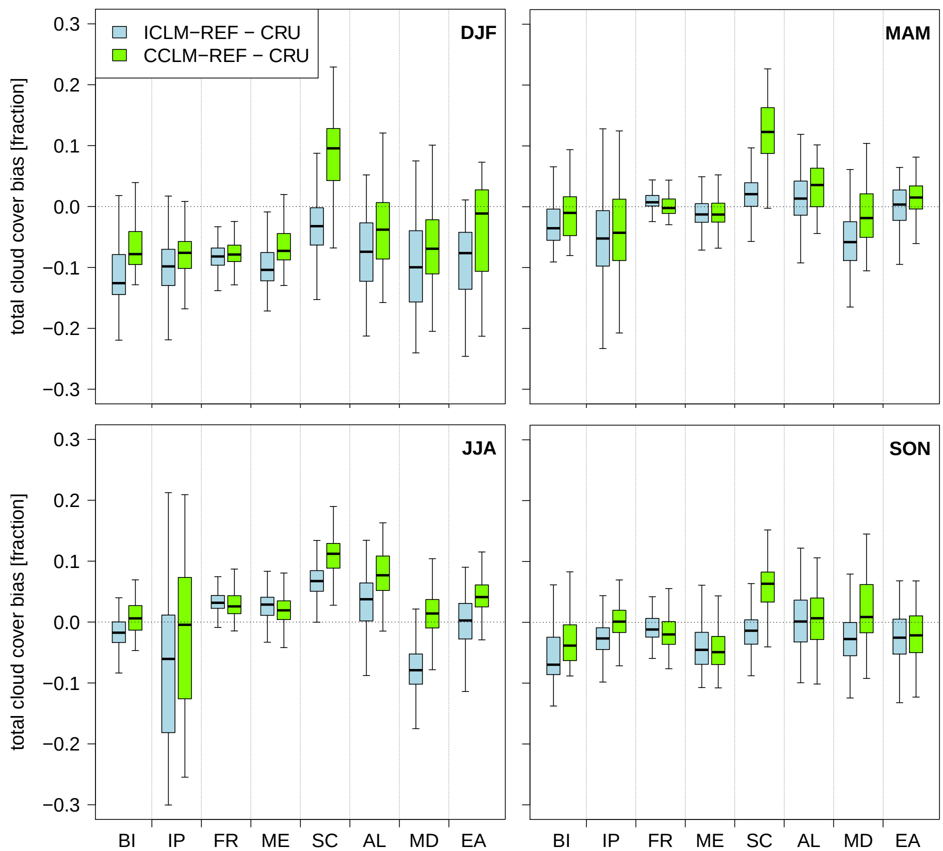

Figure 10. Same as Figure 5 but for total cloud cover.

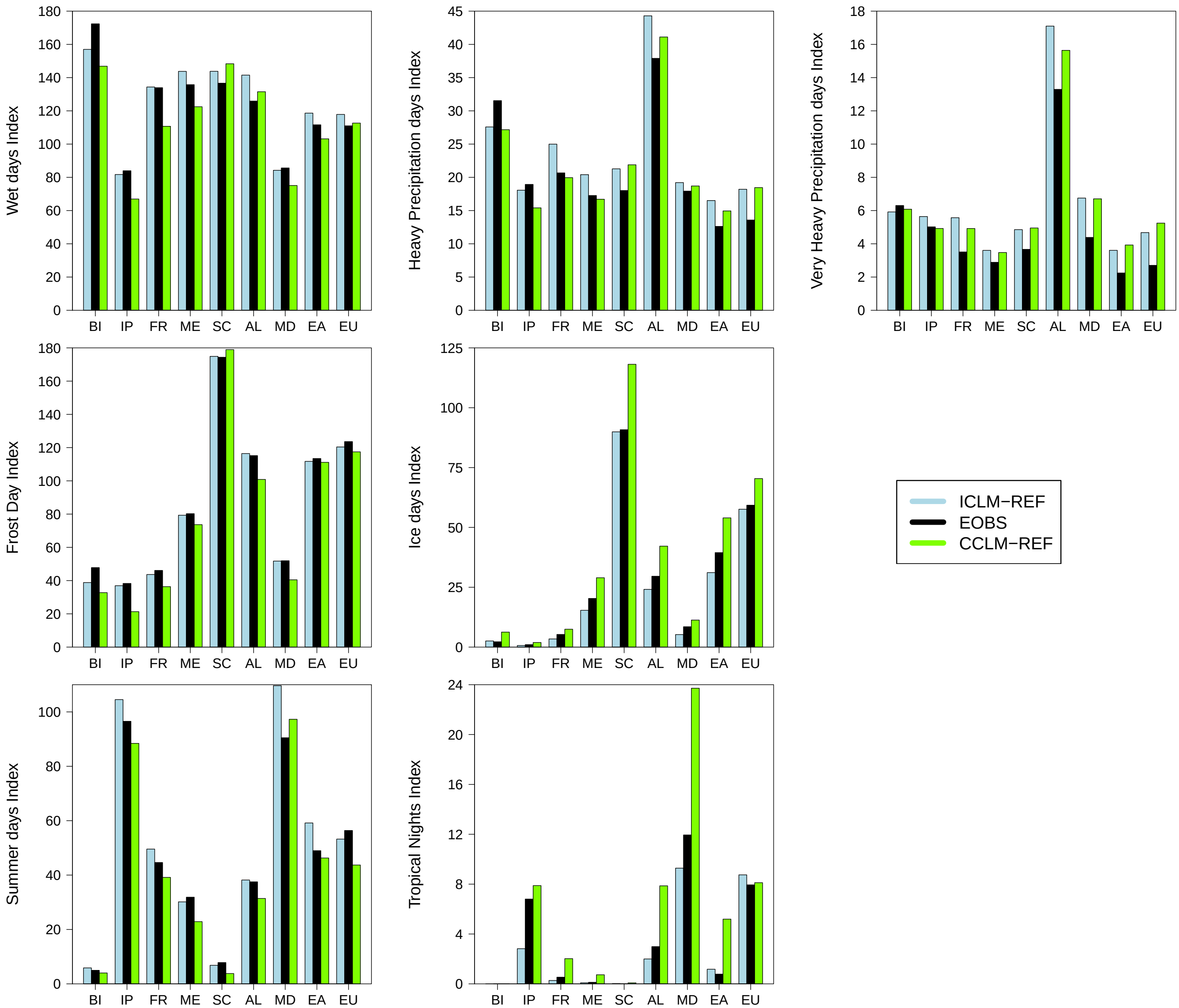

Figure 11. Area mean of climate indices [days per year] calculated from daily maximum and minimum air temperature and daily precipitation data of ICLM-REF, CCLM-REF and E-OBS for the PRUDENCE sub-regions and for the whole Europe (denoted EU) averaged over the period 1981-2000.

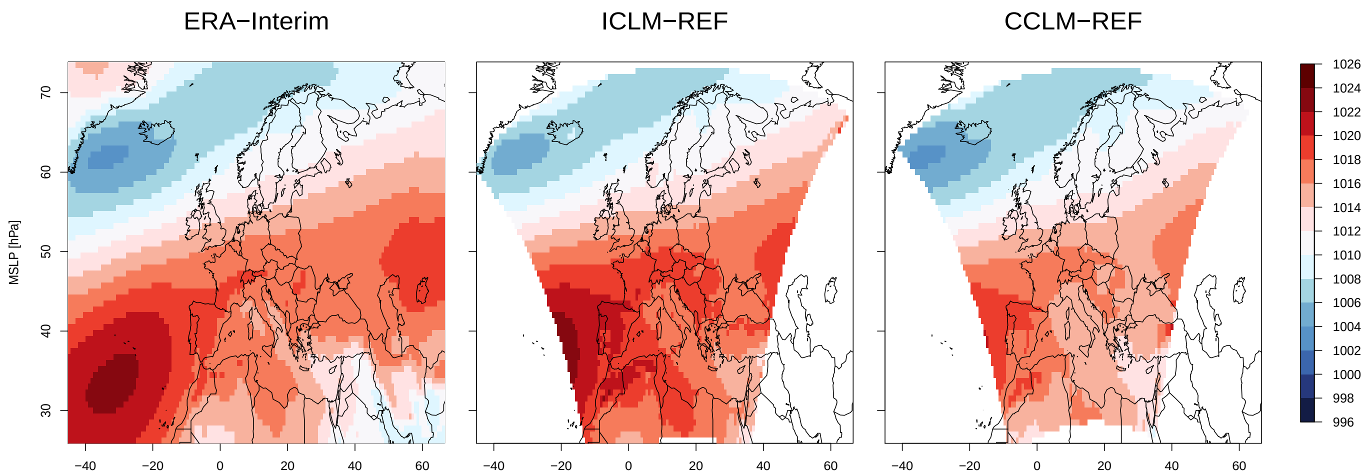

Figure 12. Annual mean sea level pressure (MSLP, in hPa) from ERA-Interim (left), ICLM-REF (middle) and CCLM-REF data (right). Data were average over the period 1981-2000.

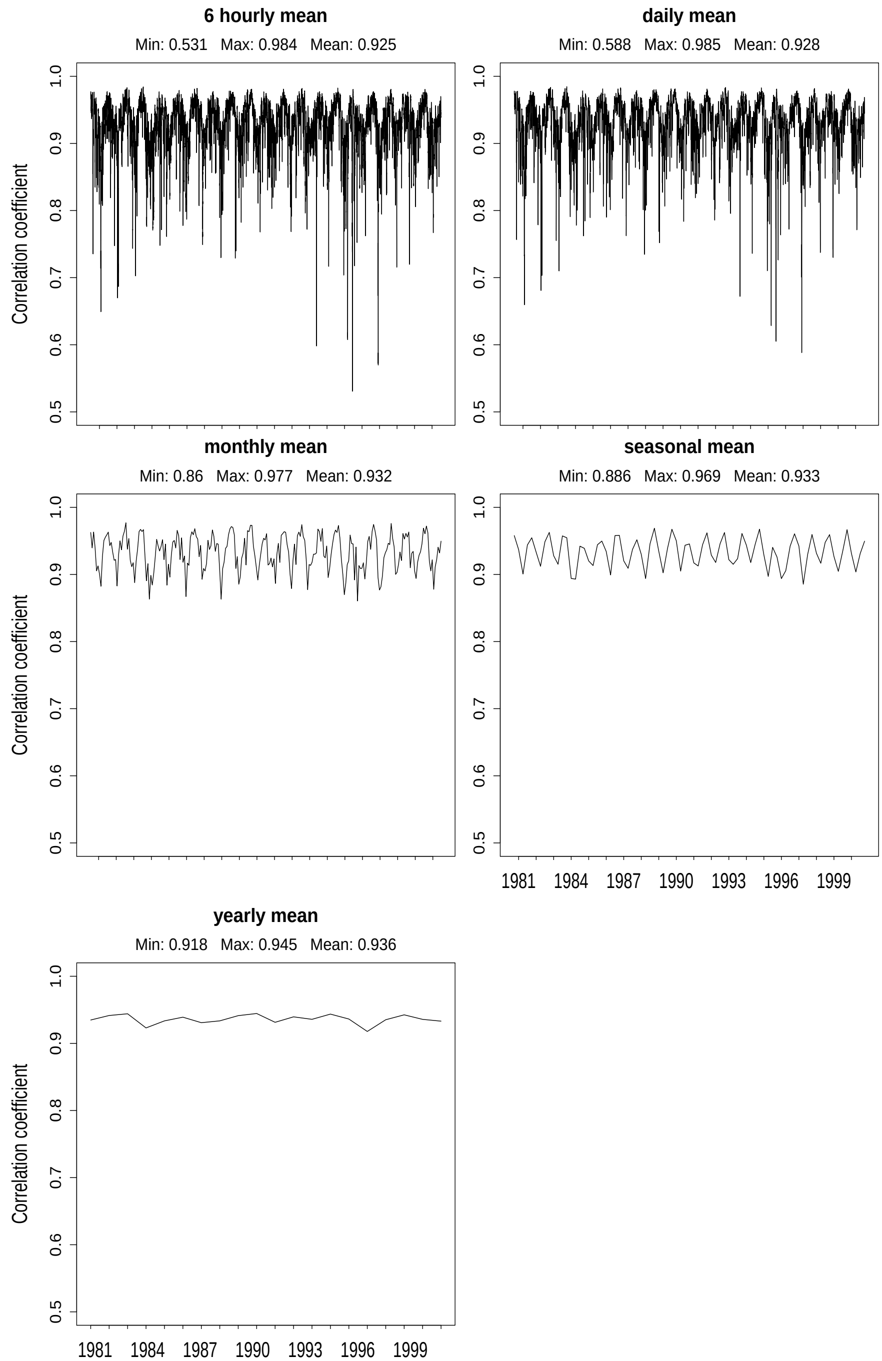

Figure 13. Spatial correlation coefficient of geopotential at 500 hPa between ICLM-REF and ERA-Interim data for various temporal scales, 6 hourly, daily, monthly, seasonal and yearly. Data were taken in the period 1981-2000.

Table 1. Total numbers of triangle cells, edges and effective grid resolution of the grids mentioned in this paper.

Grid	Domain	Number of triangle cells	Number of triangle edges	Effective grid resolution [km]
R2B8	EURO-CORDEX	286 824	430 988	9.9
R3B7	global	2 949 120	4 423 680	13.2
R3B8	Europe nested in global R3B7	659 156	989 911	6.6

Table 2. Comparison between COSMO-CLM and ICON-CLM models.

	COSMO-CLM	ICON-CLM
Planetary boundary layer scheme	Prognostic turbulent kinetic energy closure Doms et al. (2011)	Wave dissipation at critical level Orr et al. (2010)
Cumulus convection scheme	Shallow convection: Reduced Tiedtke scheme for shallow convection only (Tiedtke, 1989) Deep convection: Tiedtke (1989) oder Tiedtke-Bechtold scheme from IFS Model	Mass flux scheme with shallow, deep or mid-level convection. CAPE (convective available potential energy) based closure for deep convection. Boundary layer equilibrium closure for shallow convection. Large-scale omega (vertical velocity in pressure coordinates) based closure for mid-level convection. Tiedtke (1989); Bechtold et al. (2008)
Cloud microphysic scheme	Seifert and Beheng (2001), reduced to one-moment scheme	Single-moment scheme Doms et al. (2011); Seifert and Beheng (2001)
Radiation short and long wave	δ two-stream radiation scheme after Ritter and Geleyn (1992)	RRTM (Rapid Radiative Transfer Model) (Mlawer et al., 1997; Barker et al., 2003)
Land surface and soil scheme	TERRA-ML Doms et al. (2011)	Tiled TERRA Schrodin and Heise (2001); Schulz et al. (2016)
Coordinate system	– horizontal: rotated geographical (lat/lon) – vertical: terrain following Gal-Chen height coordinate (Gal-Chen and Somerville, 1975) and exponential height coordinate (SLEVE) according to Schär et al. (2002)	– horizontal: icosahedral grids – vertical: terrain following Gal-Chen height coordinate (Simmons and Burridge, 1981) and exponential height coordinate (SLEVE) according to Schär et al. (2002) and Leuenberger et al. (2010)

Table 3. Description of climate indices.

Index	Description	Unit
Frost days index	Number of days with minimum 2-m temperature < $0°C$	Days
Ice days index	Number of days with maximum 2-m temperature < $0°C$	Days
Summer days index	Number of days with maximum 2-m temperature > $25°C$	Days
Tropical nights index	Number of days with minimum 2-m temperature > $20°C$	Days
Wet days index	Number of days with total precipitation ≥ 1 mm	Days
Heavy precipitation days index	Number of days with total precipitation > 10 mm	Days
Very heavy precipitation days index	Number of days with total precipitation > 20 mm	Days

Table 4. RMSE and spatial STDEV ratio of averaged monthly 2-m temperature for the PRUDENCE sub-regions. Data from ICLM-REF, CCLM-REF and E-OBS from 1981 to 2000.

	RMSE [K]		Standard deviation ratio	
	ICLM-REF	CCLM-REF	ICLM-REF/EOBS	CCLM-REF/EOBS
BI	0.34	0.39	1.08	1.16
IP	0.54	0.44	0.99	0.97
FR	0.53	0.56	1.00	0.98
ME	0.56	0.73	0.99	0.98
SC	0.60	0.76	0.97	1.02
AL	0.53	0.59	1.03	1.08
MD	0.65	0.87	0.92	0.92
EA	0.87	1.03	1.10	1.12

Table 5. Same as Table 4 but for minimum 2-m temperature.

	RMSE [K]		Standard deviation ratio	
	ICLM-REF	CCLM-REF	ICLM-REF/EOBS	CCLM-REF/EOBS
BI	0.36	0.87	1.03	1.08
IP	0.81	1.15	0.96	0.88
FR	0.69	0.95	0.96	0.89
ME	0.72	1.05	1.06	0.93
SC	0.81	0.84	0.90	0.89
AL	0.75	1.06	1.13	1.19
MD	0.60	1.53	0.94	0.90
EA	0.85	1.38	1.03	1.07

Table 6. Same as Table 4 but for maximum 2-m temperature.

	RMSE [K]		Standard deviation ratio	
	ICLM-REF	CCLM-REF	ICLM-REF/EOBS	CCLM-REF/EOBS
BI	0.42	0.88	1.20	1.25
IP	0.81	0.81	1.02	1.08
FR	0.86	0.77	1.01	1.00
ME	0.76	1.01	1.06	1.09
SC	0.58	1.91	1.02	1.12
AL	0.80	1.11	0.99	1.03
MD	1.54	1.08	1.05	1.05
EA	1.29	1.19	1.18	1.12

Table 7. Same as Table 4 but for monthly accumulated total precipitation.

	RMSE [mm]		Standard deviation ratio	
	ICLM-REF	CCLM-REF	ICLM-REF/EOBS	CCLM-REF/EOBS
BI	16.78	16.03	0.80	0.80
IP	10.48	11.85	1.11	0.94
FR	16.67	14.23	1.42	1.18
ME	13.72	12.43	1.09	0.87
SC	14.77	14.88	1.15	1.08
AL	30.52	22.53	1.38	1.27
MD	17.32	12.36	1.66	1.47
EA	15.91	12.19	1.30	1.06

Table 8. Same as Table 4 but for MSLP.

	RMSE [hPa]		Standard deviation ratio	
	ICLM-REF	CCLM-REF	ICLM-REF/EOBS	CCLM-REF/EOBS
BI	1.60	0.91	1.02	0.94
IP	1.90	0.77	1.19	1.29
FR	1.71	1.12	1.38	1.11
ME	1.43	1.56	1.04	0.91
SC	1.21	0.91	0.93	0.74
AL	1.32	1.87	1.76	2.09
MD	1.68	1.79	0.74	0.73
EA	1.44	1.91	1.21	0.93

Table 9. Same as Table 4 but for total cloud cover.

	RMSE [fraction]		Standard deviation ratio	
	ICLM-REF	CCLM-REF	ICLM-REF/CRU	CCLM-REF/CRU
BI	0.08	0.06	1.00	1.00
IP	0.10	0.09	1.33	1.33
FR	0.07	0.08	1.00	0.75
ME	0.07	0.07	0.33	0.67
SC	0.06	0.10	1.33	1.67
AL	0.07	0.08	1.00	1.00
MD	0.08	0.08	0.83	0.83
EA	0.07	0.06	1.67	1.67